

YULY, ANDREA C., M.A. Children's Inferences about Gender Ambiguous People. (2019)

Directed by Janet J. Boseovski. 74 pp.

Little research has investigated children's perceptions of gender ambiguous individuals who are not part of a binary gender category. Five- to 8-year-olds were tested to determine whether appearance or conceptual information (i.e., occupation information, trait information) was prioritized when grouping a gender ambiguous target with other people. Children heard two stories with verbalized or no verbalized gender information. The target shared conceptual qualities with a character of stereotypical gender appearance (conceptual match) and appearance qualities with a character of gender ambiguous appearance (appearance match). Participants were asked which character the target should be friends with and which character's novel activity preference the target shared. Results indicated that the inclusion of verbalized gender information did not systematically change children's response patterns. However, older children prioritized trait information over appearance information, while younger children valued trait and appearance information equally. For the occupation story, children did not overwhelmingly prioritize occupation information, indicating that they might not regard occupation information as much as trait information when making predictions about gender ambiguous people. Implications include a better understanding of children's beliefs about individuals who exist outside of strict social categories, which can inform ways to combat negativity and promote acceptance of people who deviate from group-based expectations.

CHILDREN'S INFERENCES ABOUT GENDER AMBIGUOUS PEOPLE

by

Andrea C. Yuly

A Thesis Submitted to  
the Faculty of The Graduate School at  
The University of North Carolina at Greensboro  
in Partial Fulfillment  
of the Requirements for the Degree  
Master of Arts

Greensboro  
2019

Approved By

---

Committee Chair

## APPROVAL PAGE

This thesis written by ANDREA C. YULY has been approved by the following committee of the Faculty of The Graduate School at the University of North Carolina at Greensboro.

Committee Chair \_\_\_\_\_  
Dr. Janet J Boseovski

Committee Members \_\_\_\_\_  
Dr. Stuart Marcovitch

\_\_\_\_\_  
Dr. Julia L. Mendez Smith

\_\_\_\_\_  
Date of Acceptance by Committee

\_\_\_\_\_  
Date of Final Oral Examination

## ACKNOWLEDGEMENTS

I would like to thank my advisor, Janet J. Boseovski, for her mentorship and support throughout this project. I also want to thank my committee members, Drs. Stuart Marcovitch and Julia L. Mendez Smith. Also, I would also like to thank my colleagues at the DUCK Lab, especially the undergraduate research assistants that helped me with data collection (Kestyn Harris) and data coding (Nicole Fanatico). Lastly, I would like to thank my family and friends for their unconditional support.

## TABLE OF CONTENTS

	Page
LIST OF TABLES .....	vi
LIST OF FIGURES .....	vii
CHAPTER	
I. INTRODUCTION .....	1
Gender Category Biases: The Role of Labeling and Appearance.....	4
What Matters Outside of Gender Category and Appearance?.....	7
The Current Study .....	11
II. METHOD .....	15
Participants .....	15
Materials .....	15
Design.....	16
Procedure.....	16
III. RESULTS .....	24
Baseline Condition .....	24
Main Analyses.....	25
Secondary Measures.....	31
Exploratory Analyses .....	32
IV. DISCUSSION .....	35
What About Labeling? .....	42
Gender Differences for Friendship Predictions.....	44
Limitations and Future Directions.....	46
Conclusion.....	48
REFERENCES .....	50
APPENDIX A. TABLES.....	56

APPENDIX B. FIGURES .....	63
APPENDIX C. STIMULI AND VIGNETTES .....	66
APPENDIX D. MEMORY CHECK .....	69
APPENDIX E. DEPEDENT MEASURES .....	71
APPENDIX F. ESSENTIALISM MEASURE .....	72
APPENDIX G. DROODLES TASK (FROM LAGATUTTA ET AL., 2010).....	73

## LIST OF TABLES

	Page
Table 1. Descriptive Statistics and Tests Against Chance for Matching in Baseline Condition by Age Group .....	56
Table 2. Frequency of Responses for Ambiguity Check Questions for the Unlabeled Condition .....	57
Table 3. Frequency of Responses for Ambiguity Check Questions for the Labeled Condition.....	58
Table 4. Means, Standard Deviations, and Tests Against Chance for Matching in the Labeled Condition by Age Group and Story Type.....	59
Table 5. Means, Standard Deviations, and Tests Against Chance for Matching in the Unlabeled Condition by Age Group and Story Type .....	60
Table 6. Correlations Between Matching Variables and Essentialism Measure .....	61
Table 7. Correlations Between Matching Variables and Doodles Task.....	62

## LIST OF FIGURES

	Page
Figure 1. Story Information Type x Age Group Interaction.....	63
Figure 2. Story Information Type x Participant Gender Interaction.....	64
Figure 3. Participant Gender x Age Group Interaction.....	65



## CHAPTER I

### INTRODUCTION

Gender is a salient social category that is often perceived as binary (i.e., boy/girl). Beginning early in childhood, gender cues are evident through dichotomous labels, appearances, and activities (Martin, Ruble, & Szkryabło, 2002; Shutts, 2013). Gender categories are inductive, as members of a gender group are often labeled similarly and assumed to look and act alike. In fact, preschoolers assume that a boy with girl-like features shares properties with other boys, regardless of appearance (e.g., Gelman, Collman, & Maccoby, 1986). Despite associations between gender categories and specific labels, appearances, and behaviors, people can deviate from category-based expectations. Someone perceived as gender ambiguous and who does not appear to belong to a gender category might cause confusion. Those who appear gender ambiguous lack appearance, label, or behavioral cues that subsume them to a specific gender category. Gender category-based assumptions are then compromised and possibly ineffective. Given the inductive power of gender, how might children reason about a perceptually gender ambiguous person? Perceptual gender ambiguity and its implications for how children reason about category-based deviations have yet to be investigated systematically. The current study examined whether appearance, or traits and preferences, guide children's inferences about an individual who appears gender ambiguous.

Little research has explored perceptions of gender ambiguity, despite the existence of individuals in the real world who do not adhere to the binary gender categories and despite calls for investigations of individuals outside of strict social categories (Dunham & Olson, 2016). However, existent findings regarding gender-related beliefs and stereotyping provide a foundation for understanding children's beliefs about gender ambiguity. Children develop gender stereotypes in preschool that peak at about 5 years of age and reach flexibility by 9 years of age (e.g., Martin, Fabes, Evans, & Wyman, 1999; Ruble, Martin, & Berenbaum, 1998; Trautner et al., 2005). These beliefs are evident when both gender category information and conceptual information about a person are provided, such as preferences for specific activities. For example, 5-year-olds endorsed the stereotype that dolls are for girls, regardless of a girl's personal preferences (Martin, 1989). Nine-year-olds instead believed that girls do not need to like dolls. Rather, these children reported that a girl can engage in other behaviors that are similar to her preferences and stereotypical of boys. By 9 years of age, children do not prioritize gender category over a person's personal preferences and, instead, both types of information are regarded (Martin, 1989). Similarly, when provided with gender-based information (i.e., labeled as and appears gender ambiguous) and conceptual information about a perceptually gender ambiguous individual, children's beliefs about the individual might shift with age.

When reasoning about a gender ambiguous person, labels and appearance might provide sources of information. Labels and appearance cues might lead children to reason about gender ambiguity as they do for the binary gender groups: individuals labeled as

and that look gender ambiguous must hold similar preferences as and be grouped with other *gender ambiguous* people. Younger children (e.g., 5- to 6-year-olds) lack flexibility when gender expectations are defied and might infer that a gender ambiguous individual should be grouped with others who are labeled and appear gender ambiguous. Conceptual information in the form of traits or preferences for activities provide another source of reference: individuals labeled as and that look gender ambiguous, but are shy, should hold similar preferences as and be grouped with other *shy* people. Because older children (e.g., 7- to 8-year-olds) exhibit greater flexibility in their gender stereotype beliefs, they might infer that a perceptually gender ambiguous person should be grouped with others who hold similar conceptual qualities.

The current study investigated children's reasoning about gender ambiguity. Specifically, the study examined whether 5- to 8-year-olds prioritize appearance or conceptual information when asked to group a perceptually gender ambiguous person with perceptually gender ambiguous others or gender stereotypical others. This extends findings from previous research designed to compare appearance to conceptual information (e.g., Heyman & Gelman, 2000), as past studies were not done with explicit consideration of appearance ambiguity. I will begin by reviewing literature that demonstrates the inductive power of gender categories, including differences that arise through labeling and appearance manipulations. I will then review past research that compares category and appearance information to conceptual information when predictions are made about an individual. Lastly, the current study will be described in detail.

## **Gender Category Biases: The Role of Labeling and Appearance**

Children use gender categories to determine how much they like someone, what activities that person engages in, and other characteristics about the individual (Ruble et al., 1998). Labels for specific items (e.g., “This toy is for boys.”) and appearance manipulations that demonstrate lack of adherence to gender stereotypes (e.g., a boy wearing a dress) reflect children’s tendency to use gender as a basis for making inferences about another person. Investigations of whether gender ambiguity impacts the use of gender category based beliefs to make predictions about another person are scant, but existent findings provide a basis for exploration.

**Labeling.** Children show gender biases when objects are labeled *for* a gender category. When 4- to 9-year-olds were presented with gender neutral objects labeled for boys, girls, or both, they explored objects labeled for their own gender category more often and had better memory for those objects (Bradbard, Martin, Endsley, & Halverson, 1986). Labeling also dictates children’s inferences about others. Four- and 5-year-olds preferred novel toys labeled for their own gender group and extended this gender-based preference to others (Martin, Eisenbud, & Rose, 1995). Labeling individuals to a gender group induces similar biases (Shutts, Roben, & Spelke, 2013).

Labeling relates to children’s reasoning about perceptually gender ambiguous people because ambiguity might be unfamiliar. Without a gender ambiguous label, 5- and 6-year-olds might automatically categorize a gender ambiguous person as a boy or girl, grouping them with others who have a stereotypical boy or girl appearance. Such grouping might be most evident if similarities (e.g., traits or preferences) are provided

between the individual and someone of gender stereotypical appearance. However, labeling someone as *gender ambiguous* might lead 5- and 6-year-olds to predict that the person should be grouped with others who look gender ambiguous, given their reliance on gender labeling as a basis for inference (Martin et al., 1995). Seven- and 8-year-olds might look for grouping cues above and beyond labeling (Martin, 1989), resulting in a reduced tendency to group based on ambiguous appearance and label.

**Appearance.** Deviations from gender appearance expectations reveal gender biases, as children are aware of gendered features in hairstyle, clothing, and other physical markers (e.g., Halim et al., 2014). For example, when shown a boy with counter-stereotypical appearance features (e.g., long hair), 3- to 11-year-olds regarded the boy negatively (Blakemore, 2003). Violations against gender-based expectations in other domains (e.g., preferences) are not regarded as negatively: a boy with long hair received harsher judgment than a boy who liked to play with tea sets (Blakemore, 2003). Still, the negativity regarding appearance violations does not indicate the absence of gender category-based inferences about others. A boy with long hair might receive negative judgment, but that does not equate to the prediction that he should not behave similarly to, prefer the same activities as, or be grouped with other boys. Similarly, a perceptually gender ambiguous individual might receive negative judgment, but it is unclear what predictions would be made about the individual's behavior, preferences, and grouping.

Gender ambiguity is different from stereotypical appearance deviations because it denotes a lack of belonging to a dichotomous gender group, rather than holding features of the opposite gender group. Because gender ambiguity deviates from expectation,

negative judgments about gender ambiguity may be exhibited across childhood. Still, gender categories are inductive by 4 years of age, regardless of appearance deviations (Gelman et al., 1986). Four-year-olds who were shown a drawing of a child with a counter-stereotypical appearance and a gender label used the labeled gender category as a basis for prediction: a boy with long hair would like to play with trucks, rather than dolls (Gelman et al., 1986). Relatedly, children often believe that people of the same gender category share internal characteristics that link them together, also known as gender essentialism (e.g., Gelman & Taylor, 2000). Four- to 6-year-olds typically support gender essentialist endorsements, but depart from those beliefs by 9 to 10 years of age (Taylor, Rhodes, & Gelman, 2009). In parallel, the privilege granted to gender categories over other forms of information might also decline in middle childhood.

A gender ambiguous individual might be linked to other gender ambiguous individuals or be categorized to a dichotomous and inductive gender category. When labeled, the individual might be assumed to behave similarly to, prefer the same activities as, or be grouped with other gender ambiguous people. Without a label, the individual might be categorized into an existent gender category and assumed to behave similarly to, prefer the same activities as, or be grouped with other members of that gender group. Both tendencies may be typical in 5- to 6-year-olds, who often endorse gender essentialist beliefs and group-based thinking. Seven- and 8-year-olds, who depart from gender essentialism endorsements and regard gender more flexibly, may look outside of appearance and gender categories to base their predictions about perceptually gender ambiguous people, regardless of labeling. This brings to question what other

characteristics can be used to make predictions about perceptually gender ambiguous individuals.

### **What Matters Outside of Gender Category and Appearance?**

People hold characteristics beyond appearance and gender, such as traits and preferences. Traits can be used to make predictions about others (Heyman & Gelman, 2000). Likewise, activity preferences provide insight about someone's engagement in specific past times, occupations, or behaviors. If children prioritize conceptual information (i.e., traits, activity preferences) over gender, they might use it to make inferences about a person's other behaviors, preferences, and grouping. Previous studies compared conceptual information to gender category labels and appearances (Diesendruck & haLevi, 2006; Heyman & Gelman, 2000) and inform current knowledge about the interplay between gender ambiguity and conceptual information.

In the developmental literature, a key paradigm that compares conceptual information to gender labels and appearances involves two characters that share different characteristics with a third, target character. For example, in one study, children were told about 1) a target, such as a girl who is shy, 2) a character that shares gender category information with the target, such as a girl who is outgoing and likes to play zigo, and 3) a character that shares trait information with the target, such as a boy who is shy and likes to play zaber (Diesendruck & haLevi, 2006). When 5- and 6-year-olds were asked if the target liked to play the same novel activity as the gender or trait match (e.g., "Do you think she likes to play zaber like him, or zigo like her?"), traits were prioritized as often as gender (Diesendruck & haLevi, 2006). Still, collapsed across social categories,

children prioritize social categories over trait information generally (Diesendruck & haLevi, 2006). When all trio members belong to the same gender category, this paradigm can be used to compare appearance with trait information.

Results from this paradigm reflect children's use of trait over appearance information to guide their inferences about others by preschool age (e.g., Heyman & Gelman, 2000). This is in line with previous demonstrations that children use traits to predict behaviors (e.g., Liu, Gelman, & Wellman, 2007). Three- to 4-year-olds were introduced to a trio: trait information was shared between the target and one character, and appearance markers (e.g., style of hair) were shared between the target and the remaining character (Heyman & Gelman, 2000). Across age groups, children inferred that the target preferred to play the same novel activity as the trait match over the appearance match (Heyman & Gelman, 2000).

Conceptual information can also include activity preferences, which might interplay differently with gender category and appearance information. Activity information may be susceptible to gender-based biases, as children are aware of stereotypes pertaining to toys and occupations (Liben & Bigler, 2002). Children assumed that a boy who wanted to be an auto mechanic also engaged in other masculine activities and should be grouped with boys (Liben & Bigler, 2002). It is uncertain whether the activity and its gendered associations, or gender category, guided children's reasoning, as the trio paradigm has not been used to compare activities to gender information. To determine whether children prioritize activities or gender information when reasoning about others' behaviors, preferences, and groupings, gender neutral activities can be



incorporated into the trio paradigm. A target might prefer the same neutral activity as one character and share gender information with another character, allowing the influence of gender category and activity information to be teased apart.

Some comparisons demonstrate developmental changes in the privilege attributed to gender over activity information. When provided with gender category and conceptual information in the form of gender stereotypical, counter-stereotypical, or neutral toy interests, 3- to 5-year-olds prioritized gender category labels over conceptual information for making preference predictions about others (Martin, 1989). Predictions were based on a girl labeled as a girl, rather than her preference for playing with trucks (Martin, 1989). By contrast, 6- to 10-year-olds integrated category labels and conceptual information when making predictions about others. Compared to trait information, activity information interacts with gender information differently to form children's predictions about a person's interests, behaviors, and groupings.

Traits may be used more than activities to guide inferences about others. Traits are at least comparable to gender category as a basis for reasoning (e.g., Diesendruck & haLevi, 2006) and involve features that might inform behavioral predictions more than activity preferences. Activities are often disregarded by younger children when given in conjunction with category information (e.g., Martin, 1989). Comparisons of gender and appearances to traits and activities separately exist, but the role of gender ambiguity in those relationships has not been considered. A lack of gender category elicits the need to move beyond gender category based beliefs.

The trio paradigm can be adapted to determine whether conceptual information or labeled ambiguity and appearance serve as a basis for predicting how a perceptually gender ambiguous person should be grouped. A perceptually gender ambiguous target can share appearance information with one character (i.e., gender ambiguous) and conceptual information (i.e., traits, activities) with another character. Different types of conceptual information (traits, activities) must also be considered to determine if they hold similar influence over grouping. Further, the trio paradigm can be expanded by manipulating the explicit labeling of gender ambiguity. This allows the role of labeling on children's reasoning about perceptual gender ambiguity to be investigated, as a perceptually gender ambiguous person might be automatically categorized into an existing gender category or children may generally lack familiarity with gender ambiguity. Shifts across development in children's responses patterns can also be examined, given a more flexible understanding of gender stereotypes and less reliance on gender category information with age. Age differences can elucidate how response patterns may change in parallel with other developmental changes, such as those related to gender essentialism and theory of mind, or the knowledge of others' mental states (Wellman & Liu, 2004).

Children's dependence on gender categories as a source of reasoning and their rigidity about gender stereotypes declines with age, which coincides with improvements in theory of mind (Mulvey, Rizzo, & Killen, 2016). By the end of preschool, children typically demonstrate false-belief theory of mind, which means they can acknowledge others' mental states (Wellman & Liu, 2004). Interpretive theory of mind is the ability to

acknowledge that individuals can form distinct interpretations from the same experience and develops by 7 to 8 years of age (Ross, Recchia, & Carpendale, 2005). Relatedly, 6- to 9-year-olds demonstrate an overinterpretive theory of mind bias that focuses on person-specific explanations that deny the relevance of previous, common experiences between people (Lagattuta, Sayfan, & Blattman, 2010). This is particularly strong for 6- to 7-year-olds and less so in 8- to 9-year-olds (Lagattuta et al., 2010). Relatedly, children with an overinterpretive theory of mind may use conceptual information to group a perceptually gender ambiguous person. Those without an overinterpretive theory of mind may use appearance and category information to guide their grouping decisions about the target, as they may be less focused on individual-specific, conceptual information.

### **The Current Study**

For the current study, 5- to 8-year-olds were asked to group a perceptually gender ambiguous target with another perceptually gender ambiguous person or a stereotypical person that holds conceptual similarities with the target. Using the trio paradigm (e.g., Heyman & Gelman, 2000), the target and a gender stereotypical character shared conceptual information (e.g., trait or occupation preference), while the target and a gender ambiguous character shared appearance information (e.g., both do not look to belong to a gender category). To investigate potential contributors to grouping decisions about the target, the explicit labeling of gender ambiguity was manipulated and participants were exposed to multiple forms of conceptual information (i.e., trait, occupation). A baseline condition, in which all similarities were shared between the target and gender ambiguous character, was also used. If children matched the target and

gender ambiguous character, it would suggest that children understand that conceptual similarities serve as a basis for matching, even more so if supplemented with appearance similarities. Any matching between the target and the gender stereotypical character would not be based on conceptual or appearance similarities. The latter type of matching would suggest a bias for choosing the gender stereotypical character and interferes with interpretations of whether children rely on conceptual or appearance information when grouping the target.

To determine grouping decisions, children were asked to infer who the target should be friends with (Friendship Question) and what novel activities the target preferred (Novel Activities Question). Both questions were used to determine if inferences about novel activities elicit similar responses as questions about friendship. Friendship is particularly relevant to gender: preschool aged children infer more same-gender category friendships between two individuals than cross-gender friendships and exhibit same-gender friendship preferences (e.g., Martin et al., 1999). By contrast, the Novel Activities Question involves inferences about novel properties that lack associations to any particular trait, occupation, or gender group. This allows for children's grouping decisions to indicate the influence of the information given in the present study, rather than the influence of external sources (e.g., Heyman & Gelman, 2000).

Grouping decisions were also investigated in relation to parallel changes in gender essentialism beliefs and theory of mind. Gender essentialism is the notion that members of a gender category share intrinsic characteristics that link them together

(Gelman & Taylor, 2000). Children who endorse gender essentialist beliefs prioritize gender labels and category over appearance (Gelman et al., 1986), which indicates a potential connection to gender ambiguity. A scale of gender essentialism was adapted from Rhodes and Gelman (2009) to determine possible connections to grouping decisions. Grouping decisions might have also related to an overinterpretive theory of mind bias, as mentioned earlier. The Doodles task was administered to assess whether an overinterpretive theory of mind related to children's grouping decisions about a perceptually gender ambiguous person (Lagatutta et al., 2010).

Regardless of whether ambiguity was verbally denoted and labeled, children may have automatically categorized the target into a binary category. For this reason, children were also administered an Ambiguity Check at the end of the study to determine whether they maintained the target as gender ambiguous (labeled condition) and did not exclusively perceive the target as belonging to a binary gender category (unlabeled condition).

Predictions were as follows: An interaction between labeling and age was expected for the Novel Activities Question and the Friendship Question. For both questions, 5- and 6-year-olds were expected to make more conceptual matches in the unlabeled condition compared to the labeled condition, as labeling elicits category-based thinking for younger children (e.g., Martin et al., 1995). Seven- and 8-year-olds were predicted to look beyond appearance and behave more consistently across the labeled and unlabeled conditions. A main effect of story information type was also expected for each question: children were expected to make fewer conceptual matches if exposed to

occupation, rather than trait, information. As early as 3 years of age, trait-based inferences occur despite appearance similarities (Heyman & Gelman, 2000). Given that 3-year-olds show trait-based matching, and preschoolers prioritize gender category over activity information (Martin, 1989), story information type was expected to interact with age group for each question. Five- and 6-year-olds were expected to make more conceptual matches for the trait story compared to the occupation story. Seven- and 8-year-olds were expected to make conceptual matches across both stories, given their ability look beyond gender information and integrate conceptual information (Martin, 1989). Age was also expected to interact with question type: compared to the novel activities question, the friendship question was expected to elicit gender-based attitudes and therefore elicit appearance-based matches, particularly for 5- and 6-year-olds. This was not expected for 7- and 8-year-olds, who may have reached flexibility in their gender-based attitudes and beliefs.

Further, high gender essentialism endorsements were expected to positively correlate with more appearance matches in the labeled condition, given the gender ambiguity label attributed to the target and gender ambiguous character. Lastly, children with an overinterpretive theory of mind were expected to make more conceptual matches than children without the bias.

## CHAPTER II

### METHOD

#### **Participants**

One hundred six children were recruited from preschools and after-school programs in the Mid-Atlantic, along with recruitment from an existing database of families interested in contributing to developmental research. A power analysis for a three-way interaction with a mixed design was conducted with G\*power using a medium effect size of .25, an alpha of .05, and 80% power. The analysis revealed the need for a total sample size of 104 children, divided into groups of 26 for each age. Two extra children (one 5-year-old and one 8-year-old) were also tested. A sample of 49 children was recruited to complete the baseline condition.

#### **Materials**

For the main task, photographs were used to depict two separate trios of six people (see Appendix C). Every trio included a photo of a White, gender stereotypical man or woman and photos of two White, perceptually gender ambiguous people. Each trio was presented separately on a sheet of white, laminated paper.

Materials for the Essentialism Measure included a response scale (see Appendix F). A single sheet of laminated paper with a visual 5-point Likert scale of various sized circles was used to assess children's responses.

For the Doodles task, four different line drawings were used (a pig, a boot, a boat, flowers) with a corresponding occluder for each trial type. Three figures corresponding to the participants' gender and two block houses were also used (Lagattuta et al. 2010). See Appendix G.

A portable video camera and tripod were used to record each participant's session.

### **Design**

A 2 (age group: 5.0-6.9 vs. 7.0-8.9) x 2 (labeled condition: labeled or unlabeled) x 2 (story information type: trait or occupation) mixed design was used, with condition and age group as between-subject variables and story information type as the within-subject variable. Participants in the labeled condition were given one vignette with trait information and another with occupation information. Vignettes for the unlabeled condition had all gender information replaced with non-gender information. Each vignette had the gender stereotypical character matched to the participant's gender. The order of vignette presentation was counterbalanced.

For the baseline condition, a 2 (age group: 5.0-6.9 vs. 7.0-8.9) x 2 (story information type: trait or occupation) design was used, with story information type and age group as between-subject variables. Participants heard a single vignette with either trait or occupation information, counterbalanced across participants.

### **Procedure**

**Main task.** To begin, informed consent was obtained from parents and assent from the child. Upon beginning the study, participants were told the following: "Today,



I'm going to tell you about some people. Then, I'll ask you some questions about these people, but there are no right or wrong answers to the questions. Are you ready to get started?" Then, the first trio was placed directly in front of the participant and the first vignette was read aloud by the researcher.

Given that the stereotypical characters in the vignettes were gender matched to the participant, the following examples provide a description for a girl participant. All dependent measures following the vignette presentations were given identically to all participants, regardless of gender matching.

For the labeled condition, a girl participant given the occupation vignette first was told the following: "This person [point to the gender stereotypical character] is a girl. She looks like a girl. She wants to be a cook. She likes to play tibbits. We're not sure about this person [point to the gender ambiguous character]. This person doesn't look like a boy or a girl. This person wants to be an artist. This person likes to play jimjam. We're not sure about this person either [point to the gender ambiguous *target*]. This person doesn't look like a boy or a girl, like this person [point to the gender ambiguous character]. Also, this person wants to be a cook like her [point to the gender stereotypical character]." The occupations of cook and artist were chosen because both have been judged as gender neutral by children (Liben & Bigler, 2002), therefore lacking any gender-based associations that may interfere or supplement the gender information explicitly said within the vignette (i.e., is a girl, looks like a girl).

The girl participant then heard a second vignette with trait information: "This person [point to the gender stereotypical character] is a girl. She looks like a girl. She is

shy. She likes to spend time at villing. We're not sure about this person [point to the gender ambiguous character]. This person doesn't look like a boy or a girl. This person is creative. This person likes to spend time at kranoot. We're not sure about this person either [point to the gender ambiguous *target*]. This person doesn't look like a boy or a girl, like this person [point to the gender ambiguous character]. Also, this person is shy, like her [point to the gender stereotypical character].”

The novel activities listed within each vignette were the following pairs: play tibbits/jimjam, spend time at villing/kranoot. Pairs were randomly chosen to accompany the first vignette, with the remaining pair assigned to the second vignette. The pairing of each novel activity with a specific trait or occupation was also randomized.

In the unlabeled condition, all gender information (i.e., “This person is a girl. She looks like a girl.”) was replaced with non-gender information (i.e., “This person is short. This person wears small clothes.”). This provided identical stories for boys and girls, regardless of gender matching between the stereotypical character and participant gender. To match the physical uncertainty posed by gender ambiguity in the labeled condition, participants were told the following for the target and gender ambiguous character: “We're not sure if this person is tall or short. We're not sure if they wear big or small clothes.”

The following dependent measures were given to all participants in the same manner, as gender matching was only relevant to the information provided within the vignettes in the labeled condition. First, participants answered forced choice memory check questions about each character in randomized order to ensure comprehension of the

vignette (e.g., “Did I say this person was a boy, girl, or we’re not sure?”, “Did I say this person looks like a boy, girl, or we’re not sure?”, “Did I say this person wants to be a cook or an artist?”, and “Did I say this person likes to play tibbits or jimjam?”). The ordering of the novel activity information, occupation information, and trait information for each memory question was randomized (e.g., whether cook or artist was listed first). The memory question that involved novel activities was not given for the target character. Participants who failed any of the memory check questions after more than three repetitions of the information within each question were excluded from analyses, as their comprehension of the vignette was insufficient. See Appendix D.

Following the memory check, participants were asked a Novel Activities Question and Friendship Question as the two main dependent measures. For these questions, participants were asked to infer whether the target engages in the same novel activities as the conceptual or appearance match (e.g., Novel Activities Question: “Does this person [point to target] like to play tibbits like this person who is shy [point to gender stereotypical character] or does this person like to jimjam like this person who is creative [point to gender ambiguous character]?”). The ordering of the information within the Novel Activities Question was randomized (e.g., whether the conceptual or appearance match information was presented first for the question). Participants were also asked who the target should be friends with (i.e., Friendship Question: “Who should this person [point to target] be friends with?”) to determine by which dimension (appearance or conceptual match) children categorize the target (e.g., who the target is grouped with). The use of the word “should” was based from a meta-analysis by Signorella, Bigler, and

Liben (1993), which found that gender-related attitudes about individuals are best assessed with “can” or “should.” A follow-up justification question (i.e., “Why?”) was asked after the Novel Activities Question and the Friendship Question, but purely for exploratory purposes. The ordering of the Novel Activities Question and Friendship Question was counterbalanced across conditions. See Appendix E.

**Baseline condition.** For the baseline condition, all details remained the same as the main task, except that the target and ambiguous character were matched on appearance and each form of conceptual information. Children were asked to first answer the Novel Activities Question and then the Friendship Question. This was done because the focus of the task was to determine if children grasped that common conceptual information should serve as a basis for an activity inference, especially when combined with common appearance information. The Friendship Question involved inferring the target’s friendship preferences and may have encompassed more of the participant’s social judgments than the Novel Activities Question. The latter question was designed to be based solely on the information provided within the story. See Appendix E.

### **Secondary tasks.**

***Target character activity inferences.*** Participants were asked whether they believed that the target engages in a variety of gender stereotypical or neutral behaviors across ten questions. The listed activities were adapted from previous studies that investigated gender stereotypes (e.g., Boseovski, Hughes, & Miller, 2016; Liben & Bigler, 2002).

For example, children were asked: “Does this person like to play with dolls, toy trucks, or both?” The inclusion of “both” allowed for a better understanding of children’s decisions, as a forced choice question (e.g., “Does this person like to play with dolls or toy trucks?”) would not grant children as much flexibility in their responses and may have inflated results. Each question included a 1) girl stereotypical activity, boy stereotypical activity, 2) girl stereotypical activity, neutral activity, or 3) boy stereotypical activity, neutral activity. See Appendix E.

***Character liking, affiliation, and trait attribution.*** For each character, participants were asked a question to assess personal liking (e.g., “How much do you like this person - a lot, a little, or in the middle?”) and a question to measure desire for affiliation (e.g., “How much would you like to be friends with this person - a lot, a little, or in the middle?”). Then, they were asked a trait attribution (i.e., “Do you think this person is nice, mean, or not nice or mean?”) question (Boseovski & Lee, 2006). Character order was randomized. See Appendix E.

***Essentialism measure (adapted from Rhodes & Gelman, 2009).*** Participants were asked four questions regarding gender essentialism (e.g., “Knowing if a kid is a girl or a boy tells you a lot about the kid.”). Responses to each question were answered using a visual 5-point scale, using various sized circles to demonstrate agreement (i.e., smallest circle = I really disagree, next smallest = I disagree, medium sized = I don’t agree or disagree, big = I agree, largest = I really agree). Participants were trained prior to answering the questions. For the full list of questions, the visual scale, and training details, see Appendix F.

***Doodles task (Lagattuta et al., 2010).*** To begin this task, participants were introduced to two characters (Sam and Alex) matched to the participant's gender. Two block houses were also introduced, and participants were asked to assign a house to each character. A sign with the character's name and picture was then placed on the character's house. From there, participants were told that when the characters are in their homes, "they can't hear what we are saying or see what we are doing." Then, participants completed three different trial types that were randomized in their presentation: irrelevant-past, relevant-past, and distinct-pasts. For each trial type, either Sam or Alex were taken out of their house and shown a single line drawing. Thus, for each trial, either Sam or Alex was knowledgeable about the drawing's contents. The other character was naïve about the drawing's contents.

After a single character viewed a drawing, they were placed back into their house. Then, the experimenter would place an occluder over the drawing. In the irrelevant-past trial, the visible portion of the drawing allowed for a naïve observer to distinguish what the photo was (e.g., for the flower photo, the occluder window allowed one to see the flowers). For the relevant-past trial, the occluder was placed over an ambiguous portion of the drawing that did not allow a naïve observer to distinguish what the photo depicted (e.g., the curve of a boat for the boat picture). For the distinct-pasts trial, each character saw a different picture (e.g., pig or boot) that appeared similar with the occluder. The occluder was then placed over one of the drawings.

For each drawing, participants were reminded that each character either saw or did not see the picture and were asked what each character would think the drawing was

while the occluder was over it (e.g., “Sam has never seen this picture before, what will Sam think this is?”). Then, a naïve third character was introduced (e.g., “Jo did not see or hear what we were doing. Jo has never seen this picture before.”) and participants were asked to infer what (s)he would believe the drawing was (“What will Jo think this is?”). This same procedure was repeated for all trials. See Appendix G.

***Ambiguity check.*** Children were asked whether they believed the target from each vignette belonged to a specific gender category (i.e., “Do you think this person [point to target] is a boy, a girl, or you’re not sure?”). They were also asked whether they believed each target’s appearance resembled a boy, girl, or neither (i.e., “Do you think this person [point to target] looks like a boy, a girl, or you’re not sure?”), and about the target’s friendship decisions (i.e., “Is this person friends with boys, girls, or both?”). The two targets were presented in randomized order. The ordering of each question and the order of gender presentation for each question (e.g., whether boy or girl was listed first for each different question) was also randomized.

## CHAPTER III

### RESULTS

Data from two children (one 5-year-old and one 7-year-old) were excluded because they failed to pass the memory check questions. Data from three children (two 6-year-olds and one 7-year-old) were excluded due to experimenter error. Also, because the Doodles Task was added to the protocol after testing began, only 89 children completed the task. First, I will discuss results from the baseline condition, followed by descriptive statistics and the main analyses.

#### **Baseline Condition**

All conceptual matches between the target and ambiguous character, who shared appearance and conceptual features with the target, were scored as 1. Matches between the target and the stereotypical character were scored as 0. Since each child received only one story in the baseline condition, scores ranged from 0 to 1. To investigate that children did not demonstrate a systematic bias for choosing the stereotypical character, results will be first reported when collapsed by age and story information type for the Novel Activities Question and the Friendship Question. Then, results will be reported for each story information type, given that occupation information has not been used with this paradigm in previous research.

To grasp how children generally performed, regardless of age or story information type, both factors were collapsed. Children made more conceptual matches than expected



by chance for the Novel Activities Question ( $M = .90$ ,  $SD = .31$ ) and Friendship Question ( $M = .86$ ,  $SD = .35$ ),  $ps < .001$ . This was retained when analyzing responses for the Novel Activities Question within the occupation ( $M = .79$ ,  $SD = .41$ ) and trait stories (all children made conceptual matches), along with the Friendship Question within the occupation ( $M = .92$ ,  $SD = .28$ ) and trait ( $M = .80$ ,  $SD = .41$ ) stories,  $ps < .002$ . See Table 1 for matching patterns by age group.

### **Main Analyses**

For the Novel Activities Question and the Friendship Question, participants received a score of 0 for an appearance match and 1 for a conceptual match. This created a range from 0 to 1. Children's justifications for their answers (i.e., "Why?") were coded in the following way: 0 = I don't know/irrelevant, 1 = memory errors, 2 = appearance, 3 = conceptual.

**Descriptive statistics.** First, it was necessary to demonstrate that children consistently perceived the target as ambiguous in the labeled condition and did not exclusively categorize the target to a specific gender category (boy, girl) in the unlabeled condition. Therefore, frequencies were computed for each ambiguity check question. See Table 2 and Table 3. Compared to the unlabeled condition, those in the labeled condition more often reported uncertainty about the target's gender category and whether the target appeared like a boy, girl, or whether they were unsure. Also compared to the unlabeled condition, children in the labeled condition reported that the target affiliated with both boys and girls more frequently. When broken down by story information type, the target was not exclusively categorized to a specific gender category, was not said to appear like

a member of a specific gender category, and was assumed to affiliate with both boys and girls. The highest percentage of specific gender categorization was for the occupation story, as 51.9% of children rated the target as a boy.

To determine children's general matching patterns, factors (labeling condition, story information type, and age group) were collapsed. Thus, matching scores for each question were summed across stories, resulting in Total Novel Activity Matching and Total Friendship Matching scores, each ranging from 0 to 2. Children made more conceptual matches ( $M = 1.19$ ,  $SD = .73$ ) than expected by chance for the Novel Activities Questions,  $t(105) = 2.66$ ,  $p = .01$ , such that 81.1% of children made at least one conceptual match for the Novel Activities Question. Children also made more conceptual matches ( $M = 1.27$ ,  $SD = .75$ ) than expected by chance for the Friendship Question,  $t(105) = 3.76$ ,  $p < .001$ , such that 82.1% of children made at least one conceptual match for the Friendship Question. By age group, younger ( $M = 1.09$ ,  $SD = .77$ ) and older ( $M = 1.28$ ,  $SD = .69$ ) children did not differ significantly in conceptual matches made for the Novel Activities Question,  $t(104) = -1.33$ ,  $p = .19$ . Younger children's conceptual matches did not differ from chance,  $t(52) = .90$ ,  $p = .37$ , but older children made more conceptual matches than expected by chance,  $t(52) = 2.99$ ,  $p = .004$ . Differences failed to emerge between younger ( $M = 1.21$ ,  $SD = .77$ ) and older ( $M = 1.340$ ,  $SD = .73$ ) children for the Friendship Question as well,  $t(104) = -.91$ ,  $p = .37$ . Younger children's conceptual matches did not differ from chance,  $t(52) = 1.97$ ,  $p = .06$ , but older children made more conceptual matches than expected by chance,  $t(52) = 3.38$ ,  $p = .001$ . See Tables 4 and 5

for descriptive statistics for each age group by story information type for the labeled and unlabeled conditions, respectively.

**Novel activities question.** A 2 (age group: 5- and 6-year-olds vs. 7- and 8-year-olds) x 2 (labeling condition: labeled vs. unlabeled) x 2 (story information type: trait vs. occupation) binary repeated measures logistic regression analysis was completed for the Novel Activities Question using the GENMOD procedure on SAS. There was a significant story information type x age group interaction,  $\beta = 0.62$ ,  $SE = 0.30$ ,  $p = .04$ ; see Figure 1. The risk ratio obtained from the model revealed that older children were 1.85 times more likely than younger children to make a conceptual match for the trait story. Overall, children performed at chance for the trait story ( $M = .59$ ,  $SD = .50$ ),  $t(105) = 1.77$ ,  $p = .08$ . However, there was a significant difference in conceptual matches made in the trait story by older children ( $M = .70$ ,  $SD = .46$ ) and younger children ( $M = .47$ ,  $SD = .50$ ),  $\chi(1, N = 106) = 5.60$ ,  $p = .02$ . Younger children performed at chance for the trait story,  $t(52) = -.41$ ,  $p = .68$ , but older children made more conceptual matches than expected by chance,  $t(52) = 3.11$ ,  $p = .003$ . Overall, for the occupation story, children made more conceptual matches than expected by chance ( $M = .60$ ,  $SD = .49$ ),  $t(105) = 2.17$ ,  $p = .03$ . Conceptual matches made by younger ( $M = .62$ ,  $SD = .49$ ) and older children ( $M = .59$ ,  $SD = .50$ ) did not differ significantly,  $\chi(1, N = 106) = .16$ ,  $p = .69$ . Unlike the trait story, both age groups performed at chance for the occupation story ( $ps > .05$ ).

No other effects emerged as significant: there was no effect of labeling,  $\beta = -0.22$ ,  $SE = 0.22$ ,  $p = 0.31$ , and labeling did not systematically impact conceptual matching

between age groups,  $\beta = 0.06$ ,  $SE = 0.32$ ,  $p = 0.84$ . There was also no effect of age group,  $\chi(1, N = 106) = .16$ ,  $p = .69$ , or story information type,  $\beta = -0.41$ ,  $SE = 0.24$ ,  $p = 0.09$ .

For the trait story, 59.4% of children's justifications were irrelevant, 21.7% were trait based, 9.4% are appearance based, and 9.4% suggested memory errors. Cohen's Kappa was .90. For the occupation story, 60.4% of children's justifications were irrelevant, 22.6% were occupation based, 9.4% suggested memory errors, and 7.5% were appearance based. Cohen's Kappa was .81.

**Friendship question.** A 2 (age group: 5- and 6-year-olds vs. 7- and 8-year-olds) x 2 (labeling condition: labeled vs. unlabeled) x 2 (story information type: trait vs. occupation) binary repeated measures logistic regression analysis was completed for the Friendship Question using the GENMOD procedure on SAS. No effects emerged as significant ( $ps > .05$ ). Children's conceptual matches did not systemically differ between labeling conditions,  $\beta = -0.10$ ,  $SE = 0.21$ ,  $p = 0.65$ , story information type,  $\beta = -0.06$ ,  $SE = 0.20$ ,  $p = 0.76$ , and age group,  $\beta = 0.07$ ,  $SE = 0.20$ ,  $p = 0.70$ ; no interactions were significant ( $ps > .05$ ).

Although not statistically significant, younger and older children showed different response patterns between the occupation and trait story, similar to those shown for the Novel Activities Question. For the trait story, younger children ( $M = 0.60$ ,  $SD = 0.49$ ) were at chance for conceptual matches,  $t(52) = 1.53$ ,  $p = 0.13$ , but older children ( $M = 0.74$ ,  $SD = 0.45$ ) were more likely to make conceptual matches than expected by chance,  $t(52) = 3.86$ ,  $p < .001$ . However, younger and older children did not differ significantly in their conceptual matching for the trait story,  $\chi(1, N = 106) = 2.09$ ,  $p = 0.15$ . For the

occupation story, younger children ( $M = 0.62$ ,  $SD = 0.49$ ) did not differ significantly from older children ( $M = 0.64$ ,  $SD = 0.48$ ) in conceptual matching,  $\chi(1, N = 106) = 0.04$ ,  $p = 0.84$ . Younger children's conceptual matches did not differ from chance for the occupation story,  $t(52) = 1.82$ ,  $p = 0.07$ , but older children were more likely to make conceptual matches than expected by chance,  $t(52) = 2.13$ ,  $p = 0.04$ .

For the trait story, 50.0% of children's justifications were trait based, 24.5% were appearance based, 17.0% were irrelevant, and 8.5% suggested memory errors. Cohen's kappa was .78. For the occupation story, 44.3% of children's justifications were occupation based, 30.2% were appearance based, 20.8% were irrelevant, and 4.7% suggested memory errors. Cohen's Kappa was .92.

#### **Comparing the novel activities question and the friendship question.**

Predictions about differences between the Novel Activities Question and the Friendship Question did not differ by story information type. Therefore, Total Novel Activity Matching and Total Friendship Matching scores were used. A one-way MANOVA was used to compare Total Novel Activity Matching and Total Friendship Matching across age groups. There was no significant difference in Total Novel Activity Matching and Total Friendship Matching between younger and older children,  $F(2, 103) = 1.10$ ,  $p = .34$ , Wilk's  $\Lambda = .98$ , partial  $\eta^2 = .02$ .

**Supplementary analyses.** Participant matching decisions were not hypothesized to differ based on participant gender (i.e., differences between boys' matching vs. girls' matching), as no gender differences were reported in studies that used a similar methodology (i.e., Diesendruck & haLevi, 2006; Heyman & Gelman, 2000).

Accordingly, participant gender was not included in the complete binary repeated measures logistic regression models for the Novel Activities Question and the Friendship Question. For supplementary analyses, a 2 (age group: 5- and 6-year-olds vs. 7- and 8-year-olds) x 2 (labeling condition: labeled vs. unlabeled) x 2 (story information type: trait vs. occupation) x 2 (participant gender: boy vs. girl) binary repeated measures logistic regression analysis was completed for the Novel Activities Question and for the Friendship Question.

No significant effects emerged for the Novel Activities Question, except for the story information type x age group interaction discussed previously in the main analyses,  $\beta = 0.65$ ,  $SE = 0.30$ ,  $p = .03$ .

A story information type x participant gender interaction was found for the Friendship Question,  $\beta = 0.39$ ,  $SE = 0.20$ ,  $p = 0.05$ . For the trait story, boys ( $M = .60$ ,  $SD = .50$ ) and girls ( $M = .74$ ,  $SD = .44$ ) did not differ significantly in their conceptual matches,  $\chi(1, N = 106) = 2.50$ ,  $p = .11$ . Still, girls made more conceptual matches than expected by chance,  $t(53) = 3.40$ ,  $p < .001$ , but boys' matches were at chance,  $t(51) = 1.40$ ,  $p = .17$ . For the occupation story, boys ( $M = .67$ ,  $SD = .47$ ) and girls ( $M = .59$ ,  $SD = .50$ ) also did not differ significantly in their conceptual matches,  $\chi(1, N = 106) = .74$ ,  $p = .39$ . Contrary to the trait story, boys made more conceptual matches than expected by chance for the occupation story,  $t(51) = 2.64$ ,  $p = .01$ , and girls performed at chance,  $t(53) = 1.37$ ,  $p = .18$ . See Figure 2.

A significant age group x participant gender interaction emerged for the Friendship Question,  $\beta = -0.72$ ,  $SE = 0.23$ ,  $p = .002$ . There was a main effect of age

group,  $\beta = 0.46$ ,  $SE = 0.21$ ,  $p = 0.03$ , but there was no significant difference between younger children ( $M = 1.21$ ,  $SD = .77$ ) and older children ( $M = 1.34$ ,  $SD = .73$ ),  $t(104) = -.91$ ,  $p = .37$ . Rather, boys' performance drove the interaction: younger boys ( $M = 1.00$ ,  $SD = .80$ ) made significantly less conceptual matches than older boys ( $M = 1.50$ ,  $SD = .65$ ),  $t(50) = -2.48$ ,  $p = .02$ . Younger boys did not differ from chance,  $t(26) = 1.00$ ,  $p = 1.00$ , but older boys performed above chance,  $t(25) = 3.93$ ,  $p = .001$ . Conversely, younger girls ( $M = 1.41$ ,  $SD = .69$ ) and older girls ( $M = 1.19$ ,  $SD = .77$ ) did not differ significantly in conceptual matches,  $t(52) = 1.10$ ,  $p = .28$ . Conceptual matches by younger girls were above chance,  $t(26) = 3.05$ ,  $p = .01$ , but matches by older girls were not,  $t(26) = 1.22$ ,  $p = .23$ . See Figure 3.

## Secondary Measures

**Essentialism measure.** Children received a score from 0 to 4 (0 = I really disagree, 1 = I disagree, 2 = I don't agree or disagree, 3 = I agree, 4 = I really agree) for each question on the Essentialism Measure. Scores were combined across questions to create a score of children's Total Gender Essentialism Endorsement, with a range from 0 to 16. Pearson correlations between each question on the Essentialism Measure, Total Gender Essentialism Endorsement, Total Novel Activity Matching, Total Friendship Matching, and Age Group were conducted. There was a significant negative correlation between age group and the first question on the Essentialism Measure ("Being a boy or a girl is a very important part of what makes kids who they are."),  $r = -.26$ ,  $p = .007$ . There was also a significant positive correlation between age group and the fourth question on

the Essentialism Measure (“Kids are either boys or girls, there is nothing in between.”),  $r = .29, p = .003$ . Correlations were then computed separately for the labeled and unlabeled conditions, but none emerged as significant. See Table 6.

**Doodles task.** Children’s answers were scored following guidelines from Lagattuta et al. (2010). Of particular interest were children’s responses to person-specific experiences (i.e., Distinct-Pasts task), but analyses were conducted for each task (i.e., Relevant-Past, Irrelevant-Past, Distinct-Pasts). Correlations between children’s complete scores from each Doodles Task, total performance across all Doodles Tasks, Total Novel Activity Matching, Total Friendship Matching, and Age Group were not significant. See Table 7.

### **Exploratory Analyses**

All the following measures were included for exploratory purposes, as they have not been supplemented with the trio paradigm in previous research. These measures were expected to reveal more about children’s attitudes and beliefs about the characters.

**Target character activity inferences.** Children received a score of 0 if they inferred the target preferred a boy stereotypical activity, a score of 1 if they inferred the target preferred a neutral activity or both girl and boy stereotypical activities, and a score of 2 if they inferred the target preferred a girl stereotypical activity. Summed across all questions, scores ranged from 0 to 18. There was no significant difference between activity inferences in the occupation ( $M = 7.82, SD = 3.34$ ) and the trait stories ( $M = 7.81, SD = 3.43$ ),  $t(103) = -.04, p = .97$ . Scores were then combined across stories to create a Total Activity Inference score. Combined scores ranged from 0 to 36. No significant



correlations emerged between Total Activity Inference scores and the Novel Activities Question or the Friendship Question. With age group included, there was a significant negative correlation between age group and Total Activity Inferences,  $r = -.26, p = .01$ . To determine if the inclusion of explicit gender information shifted responses to this measure, labeling conditions were compared. No effects emerged significant when comparing Total Activity Inference scores across the labeled ( $M = 15.83, SD = 4.75$ ) and unlabeled ( $M = 15.44, SD = 5.46$ ) conditions,  $t(102) = -.38, p = .70$ .

**Character liking and affiliation.** Liking and affiliation was measured with the following scale for each character: 0 = a little, 1 = in the middle, 2 = a lot. There was no significant difference for liking and affiliation scores between stories ( $ps > .05$ ). Scores were summed to create a Total Liking and Total Affiliation score for each character, creating a range from 0 - 4. No significant correlations emerged between each grouping question and Total Liking or Total Affiliation. There was also no significant correlations with story information type, labeling condition, or age group ( $ps > .05$ ). Given the 0 (a little) to 4 (a lot) range, children were neutral in their reported desire for affiliation with the target ( $M = 2.24, SD = 1.16$ ) and liking for the target ( $M = 1.96, SD = 1.13$ ).

Younger children ( $M = 2.02, SD = 1.25$ ) and older children ( $M = 1.91, SD = 1.01$ ) did not differ significantly in their liking for the target,  $F(1, 104) = .27, p = .61$ . Younger children ( $M = 2.76, SD = 1.16$ ) and older children ( $M = 2.43, SD = 1.19$ ) also did not differ significantly in their liking for the stereotypical character,  $F(1, 104) = 1.99, p = .16$ . Additionally, younger children ( $M = 2.26, SD = 1.27$ ) and older children ( $M = 2.21, SD = 1.04$ ) did not differ significantly in reported affiliation with the target  $F(1, 104) = .06, p =$

.80. Younger children ( $M = 2.89$ ,  $SD = 1.10$ ) and older children ( $M = 2.57$ ,  $SD = 1.19$ ) also did not differ significantly in reported affiliation with the stereotypical character,  $F(1, 104) = 2.08$ ,  $p = .15$ .

**Character trait attributions.** For scoring, 0 = mean, 1 = not nice or mean, 2 = nice. Character attributions did not differ significantly between story information type ( $ps > .05$ ). Scores were then summed across stories to create a Total Trait Attribution score for each character, creating a range from 0 – 4. Correlations between each character's Total Trait Attribution score and the Novel Activities Question, along with the Friendship Question, were not significant. Correlations between Total Trait Attribution scores and age group were also not significant. Still, it is important to note that children were positive ( $M = 3.32$ ,  $SD = 1.02$ ) in their trait attributions for the target, given the 0 (mean) to 4 (nice) scale.

Also, younger children ( $M = 2.27$ ,  $SD = 1.11$ ) and older children ( $M = 3.38$ ,  $SD = .95$ ) did not differ significantly in their trait attributions for the target,  $F(1, 103) = .29$ ,  $p = .59$ . Relatedly, younger children ( $M = 3.34$ ,  $SD = 1.07$ ) and older children ( $M = 3.51$ ,  $SD = .89$ ) did not differ significantly in their trait attributions for the stereotypical character,  $F(1, 104) = .79$ ,  $p = .38$ .

## CHAPTER IV

### DISCUSSION

The present study is one of the first to assess what type of information children use to reason about gender ambiguous individuals. Generally, older children used conceptual information to group the target (i.e., predict the gender ambiguous target's novel activity preferences and friendships), while younger children did not show consistent reliance on conceptual or appearance information. The type of conceptual information (i.e., trait or occupation) provided within each story also changed children's responses. For the trait story, older children used trait similarities to guide their grouping decisions for the target, but younger children were as likely to use trait or appearance information. For the occupation story, children mostly used occupation and appearance information to inform their grouping decisions for the target and this was evident across development. Children's reasoning abilities and their lack of familiarity with gender ambiguity might explain why they exhibited distinct responses across age and story information type.

Only older children consistently made trait-based novel activity and friendship inferences about the target, suggesting improved trait reasoning with age. Prior research reveals that children exhibit improvements in trait reasoning across childhood. By 7 years of age, children use traits to describe others (e.g., Livesley & Bromley, 1973) and can successfully use traits in categorical and dimensional terms (Gonzalez, Zosuls, & Ruble,

2010). Perhaps the ability to describe a person with traits and consider traits dimensionally entails a sophisticated understanding of traits, such as greater knowledge about the behavioral implications of a trait. Consequently, traits might be regarded as a strong basis for making inferences about unfamiliar others (i.e., the target), at least compared to superficial, appearance-based cues that might not entail as much information.

In addition to a complex understanding of trait information, older children's ability to regard trait information over appearance information indicates a departure from superficial, perceptual-based reasoning. Older children value perceptual information when it is the only information provided within a task, but, unlike younger children, switch to other sources of information (e.g., behaviors) when those are provided in conjunction with perceptual information (e.g., Hoffner & Cantor, 1985). This would suggest that if only appearance information was used in the present study, older children might have made appearance-based grouping decisions. However, it is critical to note that older children's tendency to overlook superficial cues (e.g., appearance) in the presence of more telling information (e.g., traits) was limited to the trait story.

Younger children used perceptual information and trait information similarly to infer the target's novel activity preferences and friendships, implying a consideration for perceptual information that is not always captured through investigations with characters of traditional appearance. In contexts outside of perceptual gender ambiguity, younger children view traits as stable and predictive, although they are biased in some ways, such as by often making positive trait attributions (Boseovski, 2010; Heyman & Gelman,

2000; Liu et al., 2007). Perceptual gender ambiguity might have triggered a focus on perceptual information that competed with the value often placed on relevant trait information. Some prior research is consistent with the idea that younger children engage in appearance-based reasoning (e.g., Hoffner & Cantor, 1985; Rholes & Ruble, 1984). For example, younger children are influenced by the strength of appearance information that is simultaneously provided with trait information (Gonzalez et al., 2010) and sometimes value appearance over trait-based behavioral information to make predictions about others (e.g., Hoffner & Cantor, 1985). Past findings suggest that appearance-based reasoning results from children's cognitive abilities. A similar explanation applies to younger children's inconsistent friendship and novel activity inferences about the target.

Specifically, younger children engage in preoperational thinking, which then results in a weakened ability to conceptualize properties about a person as enduring (Rholes & Ruble, 1984). This subsequently elicits a failure to see those properties as causal and therefore as proper sources of inference. However, it is important to integrate these suggestions with research that shows sophisticated trait reasoning in younger children (e.g. Heyman & Gelman, 2000). Younger children's inconsistent predictions about the target might entail that they failed to characterize traits as more persistent and telling of the target than appearance information. In other words, younger children saw appearance *and* trait information as stable and predictive in the context of perceptual gender ambiguity. As further evidence of this, 5-year-olds use trait and gender categories comparably when making predictions about another person (Diesendruck & haLevi, 2006). Consequently, diminished reliance on trait information might extend to contexts in

which general gender information (e.g., gender category) explicitly competes with trait information as a source of making decisions about another person.

By contrast, nuanced patterns emerged for the occupation story that require separate explanations for the Novel Activities Question and Friendship Question. First, the Novel Activities Question and Friendship Question were not significantly different, demonstrating that gender ambiguity generally diminished the presence of gender in-group biases typically found in children's friendship decisions (e.g., Maccoby, 1990; Maccoby & Jacklin, 1987). However, there were distinct trends for each question within the occupation story: across age groups, children used both occupation and appearance information to inform their novel activity predictions for the target, but only older children used occupation information consistently to guide their friendship predictions. Older children's friendship predictions hint at their increasingly flexible views of gender in middle childhood (e.g., Martin, 1989). Older children did not regard gender as a sole indicator of friendship (e.g., girls should only be friends with other girls), which led them to predict that the target could be friends with the stereotypical character. Thus, flexible gender attitudes in middle childhood, as described by previous studies (e.g., Martin, 1989), extend to perceptual gender ambiguity. However, this flexibility was limited to the Friendship Question. Children's limited understanding of occupations might provide insight into their inconsistent decisions about the target for the Novel Activities Question across development.

Generally, children have some knowledge about occupations. For example, they can associate occupations to specific gender categories (e.g., Weisgram, Bigler, & Liben,

2010). Even more broadly, they associate activities to specific gender groups (see Liben & Bigler, 2002 for review). However, associations do not entail that children have sufficient knowledge that would allow for occupation information to inform predictions about another person consistently. Relatedly, 5-year-olds exhibit a general understanding of disciplines but fail to understand the complexities and relations underlying to those disciplines (Danovitch & Keil, 2004). By extension, children might understand what an occupation generally entails, but not other characteristics of an occupation, such as the kind of commitment, education, or other more complex and non-obvious characteristics often implied for adults.

For example, when children were told that the target wanted to be an artist, perhaps they perceived the target as someone who makes art and is knowledgeable about art. General knowledge about occupations might have been enough to suggest similar activities and behaviors (e.g., an artist paints) between two people with shared occupation preferences. Children's friendships are often based on play patterns (Maccoby, 1990), indicating that they likely use activities to inform friendship predictions. The activities associated with occupations could have therefore led older children to use occupations to inform their friendship predictions for the target. It could have also led younger children to use occupation information more than trait information for the Friendship Question.

A simple understanding of occupations, coupled with a lack of familiarity with novel activities, might have led to the inconsistent use of appearance and occupation information to inform novel activity predictions across age. To continue the artist example above, children might have failed to acknowledge non-obvious features, such as

the amount of practice needed to become an artist, which routes lead to sufficient knowledge about art, or other interests and characteristics that might be common among artists. Consequently, occupation information was not indicative of much information for children, especially unfamiliar novel activity information. In fact, children engage in property-based reasoning over appearance-based reasoning when given familiar properties over unfamiliar properties (Gelman et al., 1986), entailing the role of knowledge about a property (e.g., occupations) in the subsequent and successful use of property information. Nevertheless, children recalled the occupations presented in the stories, suggesting that limited knowledge about occupations did not hinder their understanding of the story, but only the application of that information to guide their novel activity predictions for the target.

A simple understanding of occupations cannot fully explain the patterns exhibited in children's novel activity predictions. When familiar preferences (i.e., colors) were compared to appearance information, preschoolers used preferences as much as they used appearance to guide their novel activity predictions about another person (Heyman & Gelman, 2000, Experiment 3). This suggests that children might not consider preferences when asked to make inferences about others. Still, color preferences do not necessarily equate to occupation preferences, given the breadth of information occupations entail. Regardless, perhaps other factors, such as perceptual gender ambiguity, also contributed to children's inconsistent novel activity predictions for the occupation story.

In traditionally gendered contexts, children demonstrate abundant knowledge about gender (e.g., Martin et al., 2002) and show consistent patterns in their use of (or



lack thereof) activity preferences to inform their decisions about others (e.g., Martin, 1989). Older children value activity preferences over gender category, while younger children do the opposite (Martin, 1989). Thus, there is at least consistency in children's response patterns (Martin, 1989), which was missing when children were asked to infer novel activity preferences for the target. In contrast to contexts regarding known activities and the traditional gender categories, participants could not use their existing gender knowledge and experience to supplement their decisions about the target. Akin to the perceptual focus that younger kids exhibited in the trait story, a focus by older and younger children on untraditional and unfamiliar perceptual gender ambiguity during the occupation story might have brought forth inconsistent novel activity predictions.

A focus on the target's perceptual information might also explain why overinterpretive theory of mind did not relate to children's novel activity and friendship decisions about the target across development. If children focused on appearance information, rather than focusing on the target's individuating factors (an overinterpretive theory of mind; Lagattuta et al., 2010), an overinterpretive theory of mind should not be associated with predictions made about the target.

Also, reported liking, desire for affiliation, and trait attributions made toward the target did not correlate with children's friendship and novel activity predictions, regardless of age. This suggests that children's own attitudes about the target did not systematically determine whether predictions about the target were primarily appearance or conceptually based. Younger and older children reported similar liking and desire for affiliation with the target, suggesting that age was not related to changes in reported

attitudes toward the target. Children also made positive trait attributions toward the target, providing further evidence for the positivity bias reported in children through middle childhood (Boseovski, 2010). The positivity bias was reflected similarly in both older and younger children's trait attributions for the target. Since children were not overwhelmingly negative, these results contrast previous findings that reveal negative attitudes across childhood toward individuals who depart from gender stereotypical expectations (Blakemore, 2003). This provides further evidence that perceptual gender ambiguity elicited responses that do not parallel trends found in more traditionally gendered contexts.

### **What About Labeling?**

Children's lack of familiarity with perceptual gender ambiguity can also indicate why explicit gender ambiguous labels did not impact children's novel activity or friendship predictions about the target across age. In other words, the target's labeled gender ambiguity cannot change children's grouping decisions about the target if gender ambiguity is unfamiliar. The lack of labeling effect was expected for older children, but not due to a lack of familiarity with gender ambiguity. Rather, older children show flexibility in their beliefs about gender (e.g., Serbin & Sprafkin, 1986). Consequently, they were expected to go beyond labels and use other information (i.e., trait, occupation) to guide their inferences about the target. However, with increasing age, children were more likely to support the binary notion that people can only be boys or girls on the Essentialism Measure. This implies that older children's flexibility with gender beliefs

might be limited to attitudes and sources valued for prediction, rather than beliefs that individuals might exist outside the gender binary.

In contrast, younger children were expected to make fewer conceptual matches when provided with gender ambiguity labels, given their in-group biases for objects and people that are given a gender label (e.g., Bradbard et al., 1986; Martin et al., 1995). To further support this, with decreasing age, children judged gender as more important to what makes people who they are on the Essentialism Measure, further exhibiting their attention to gender category labels. The unfamiliarity of gender ambiguity labels in the present study might have made labels irrelevant when making inferences about the target, suggesting that younger children lacked essentialist beliefs that all individuals with a gender ambiguous label hold intrinsic similarities outside of appearance. In other words, younger children's gender essentialist beliefs about the traditional gender categories (e.g., Rhodes & Gelman, 2009) did not extend to gender ambiguous people. Additionally, children overlook labels when considered arbitrary and irrelevant to the task (Heyman & Gelman, 2000, Experiment 2). It is possible that children's predictions for the target would exhibit a labeling effect in more gendered contexts. For example, children could be asked whether the target engages in activities that are stereotypically associated with girls or boys. The target character activity inferences measure did this, yet children's responses did not shift in the presence of a gender ambiguity label. Nevertheless, gender biases arose unexpectedly in the contexts described below.

## **Gender Differences for Friendship Predictions**

Gender differences were exhibited in children's friendship predictions for the target, which changed between story information type and exhibit general differences in what boys and girls consider a meaningful basis for friendship. These analyses were supplementary, as no gender differences were reported in studies using a similar paradigm (e.g., Diesendruck & haLevi, 2006) and therefore participant gender was not hypothesized to differentially impact what information children used to make inferences about the target. Boys used occupation information to guide their decisions about the target's friendships, but used trait and appearance information inconsistently for the same question, suggesting that boys consider activities as more indicative of friendship than trait information. For example, since artists engage in the same activities (e.g., painting, drawing), artists might be more likely to be friends than two people with similar appearance features or traits. Consistent with this idea, friendships between boys are described as primarily activity-based (Maccoby, 1990). By contrast, girls used trait information to guide their friendship decisions for the target, but used appearance and occupation information comparably. Past findings describe friendships between girls with intimate characteristics (Maccoby, 1990). Perhaps girls perceived trait information as more intimate to the target than appearance or occupation preferences. Still, it is important to note that these gender differences were not uniform across development.

Children's beliefs about friendship competed with age-related improvements in flexibility regarding gender attitudes and biases. Compared to older boys, younger boys were less likely to use conceptual information to inform how they grouped the target for

the Friendship Question, which indicates categorical based reasoning. Regardless of trait or occupation information, younger boys might have generally believed that the target should not be friends with the gender stereotypical character, as the gender stereotypical character should only be friends with people of the same gender category. This follows past research with the traditional gender categories: younger children use gender labels and categories to make inferences about friendship (e.g., Martin & Little, 1990; Shutts et al., 2013), but older children exhibit flexibility that goes beyond categorical based thinking in their predictions about others (e.g., Martin, 1989).

Conversely, older girls made fewer conceptual matches than younger girls, suggesting that flexible gender beliefs cannot fully account for age differences in children's friendship predictions for the target. Given that girls view friendship with greater intimacy than boys (Maccoby, 1990), it follows that conceptual (whether trait or occupation) information would generally be prioritized over appearance information. At the same time, friendship groups tend to become more homogenous as children grow older and some research suggests that this is seen more in girls than in boys (Maccoby, 1990; Powlishta, 1995; Yee & Brown, 1994). Younger girls might have relied on the intimate information provided by conceptual information, at least compared to appearance information, but older girls might have *also* relied on their in-group biases that indicate friendship groups should be homogeneous and therefore the target should be friends with other gender ambiguous people and/or not a traditional gendered character. In sum, children's inferences about the target's friendships suggest underlying gender

differences in views of friendship, supplemented by age-related changes in flexibility with gender beliefs.

### **Limitations and Future Directions**

Gender categorization is an automatic process by 4 years of age (Weisman, Johnson, & Shutts, 2015), rendering it difficult to accept perceptual gender ambiguity. This automaticity is especially relevant to younger children, as they often overlook appearance features and place great value on categorical information (e.g., Gelman et al., 1986). To prevent automatic gender categorization, future studies must consider removing all forms of gender cues from stimuli. For example, silhouettes can be used that lack gender-related information and require experimenters to explicitly label gender categories or gender ambiguity. If these changes replicate the present findings, it would imply that the context of gender ambiguity changes previously established patterns pertinent to information that children use to make inferences about other people.

Additionally, children's justifications for their grouping decisions were primarily irrelevant for the Novel Activities Question, suggesting an effect of verbal limitations compounded with a lack of familiarity with novel activities. Children might have known to make an appearance or conceptually based prediction, but were unable to explain their decision verbally in an unfamiliar context (i.e., novel activities). By contrast, the Friendship Question elicited a more familiar context, as evidenced by children's increased tendency to make appearance or conceptually based justifications. Therefore, children successfully applied their prior knowledge that similarities (e.g., "both characters want to be cooks") can serve as a foundation for friendship. By contrast, novel

activities were unfamiliar and children might not have known to rely on similarities to inform their predictions for the Novel Activities Question.

To dampen the effects of verbal limitations and familiarity with novel activities, future studies should replace the open-ended justification question with a forced-choice justification question. This limits responses to appearance or conceptual reasoning. If this leads children's justifications to more closely follow their grouping decisions, justifications could provide further evidence for what information children use to make predictions about gender ambiguous individuals.

It is also critical to include additional secondary measures to capture parallel developmental processes that explain children's inferences about the target. If the perceptual novelty of the target heavily interferes with and impacts children's inferences about the target, better cognitive abilities, such as flexibility and inhibitory control, might elicit more conceptual matches over appearance matches. These abilities might be associated with less gender category-based thinking and gender biases, given that such skills could indicate the ability to overcome strong and predominant notions about gender (e.g., everyone is either a boy or a girl; girls should be friends with other girls). Those with better flexibility and inhibitory control might be more likely to group the target and gender stereotypical character together. Gender biases might be more difficult to overcome for those with less flexibility and inhibitory control, such that the target is *not* be grouped with people of the traditional gender categories due to the target's departure from gender category expectations.

Lastly, the disparate responses between the trait and occupation stories also warrant further investigation. This was not limited to the main task, as results from the baseline condition indicated more frequent appearance matches for the occupation story than for the trait story. In general, occupations and other preferences might not be regarded as better sources of prediction than trait information. Therefore, it is important to question whether occupations would overpower appearance and gender category information with only members of the traditional gender categories. Including occupation information within a trio of individuals belonging to the traditional gender categories (e.g., the target and member of the opposite gender category share occupation preferences, but the target and the remaining character share appearance and gender category) can indicate the value children place on occupation and general preference information. If occupations and preferences still do not impact children's responses, then perhaps the trio paradigm cannot capture the activity-based predictions children display through other paradigms in the gender literature.

## **Conclusion**

In sum, only older children used trait information over appearance information to make inferences about perceptually gender ambiguous people. Across age groups, children mostly used occupation information and appearance information similarly to guide their predictions about perceptually gender ambiguous individuals. Given that past findings (e.g., Gelman et al., 1986) which demonstrate the inductive power of gender do not fully account for these effects, it is necessary to expand traditional frameworks with consideration for perceptual gender ambiguity. Otherwise, developmental theories and



knowledge will continue to lack full consideration for individuals that exist in the real world. Given the experiences gender ambiguous individuals might endure in their lives due to their lack of adherence to the traditional gender categories, the literature can provide empirically based knowledge that promotes the idea of gender ambiguity in childhood and does not further marginalize individuals that do not adhere to strict gender categories.

## REFERENCES

- Bigler, R. S., & Liben, L. S. (2006). A developmental intergroup theory of social stereotypes and prejudice. In R.V. Kail (Ed.), *Advances in child development and behavior* (pp. 39-89). San Diego: Elsevier.
- Blakemore, J. E. O. (2003). Children's beliefs about violating gender norms: Boys shouldn't look like girls, and girls shouldn't act like boys. *Sex Roles*, 48(9-10), 411-419.
- Boseovski, J. J. (2010). Evidence for “rose-colored glasses”: An examination of the positivity bias in young children’s personality judgments. *Child Development Perspectives*, 4(3), 212-218.
- Boseovski, J. J., Hughes, C., & Miller, S. E. (2016). Expertise in unexpected places: Children’s acceptance of information from gender counter-stereotypical experts. *Journal of Experimental Child Psychology*, 141, 161-176.
- Boseovski, J. J., & Lee, K. (2006). Children's use of frequency information for trait categorization and behavioral prediction. *Developmental Psychology*, 42(3), 500-513.
- Bradbard, M. R., Martin, C. L., Endsley, R. C., & Halverson, C. F. (1986). Influence of sex stereotypes on children's exploration and memory: A competence versus performance distinction. *Developmental Psychology*, 22(4), 481-486.

- Danovitch, J. H., & Keil, F. C. (2004). Should you ask a fisherman or a biologist?: Developmental shifts in ways of clustering knowledge. *Child Development, 75*(3), 918-931.
- Diesendruck, G., & haLevi, H. (2006). The Role of Language, Appearance, and Culture in Children's Social Category-Based Induction. *Child Development, 77*(3), 539-553.
- Dunham, Y., & Olson, K. R. (2016). Beyond discrete categories: Studying multiracial, intersex, and transgender children will strengthen basic developmental science. *Journal of Cognition and Development, 17*(4), 642-665.
- Gelman, S. A., Collman, P., & Maccoby, E. E. (1986). Inferring properties from categories versus inferring categories from properties: The case of gender. *Child Development, 57*(2), 396-404.
- Gelman, S. A., & Taylor, M. G. (2000). Gender essentialism in cognitive development. In P. H. Miller & E. K. Scholnick (Eds.), *Developmental psychology through the lenses of feminist theories* (pp. 169-190). New York: Routledge.
- Gonzalez, C. M., Zosuls, K. M., & Ruble, D. N. (2010). Traits as dimensions or categories? Developmental change in the understanding of trait terms. *Developmental Psychology, 46*(5), 1078 - 1088.
- Halim, M. L., Ruble, D. N., Tamis-LeMonda, C. S., Zosuls, K. M., Lurye, L. E., & Greulich, F. K. (2014). Pink frilly dresses and the avoidance of all things “girly”: Children’s appearance rigidity and cognitive theories of gender development. *Developmental Psychology, 50*(4), 1091-1101.

- Heyman, G. D., & Gelman, S. A. (1998). Young children use motive information to make trait inferences. *Developmental Psychology*, 34(2), 310-321.
- Heyman, G. D., & Gelman, S. A. (2000). Preschool children's use of trait labels to make inductive inferences. *Journal of Experimental Child Psychology*, 77(1), 1-19.
- Hoffner, C., & Cantor, J. (1985). Developmental differences in responses to a television character's appearance and behavior. *Developmental Psychology*, 21(6), 1065-1074.
- Lagattuta, K. H., Sayfan, L., & Blattman, A. J. (2010). Forgetting common ground: Six- to seven-year-olds have an overinterpretive theory of mind. *Developmental Psychology*, 46(6), 1417-1432.
- Liben, L. S., & Bigler, R. S. (2002). *Monographs of the Society for Research in Child Development*, 67(2), 1-187.
- Liu, D., Gelman, S. A., & Wellman, H. M. (2007). Components of Young Children's Trait Understanding: Behavior-to-Trait Inferences and Trait-to-Behavior Predictions. *Child Development*, 78(5), 1543-1558.
- Livesley, W. J., & Bromley, D. B. (1973). Person perception in childhood and adolescence. London: Wiley.
- Maccoby, E. E. (1990). Gender and relationships: A developmental account. *American Psychologist*, 45(4), 513 – 520.
- Maccoby, E. E., & Jacklin, C. N. (1987). Gender segregation in childhood. In H. W. Reece (Ed.), *Advances in Child Development and Behavior* (pp. 239 – 287), New York: Academic Press.

- Martin, C. L. (1989). Children's use of gender-related information in making social judgments. *Developmental Psychology*, 25(1), 80 - 88.
- Martin, C. L., Eisenbud, L., & Rose, H. (1995). Children's Gender-Based Reasoning about Toys. *Child Development*, 66(5), 1453-1471.
- Martin, C. L., Fabes, R. A., Evans, S. M., & Wyman, H. (1999). Social cognition on the playground: Children's beliefs about playing with girls versus boys and their relations to sex segregated play. *Journal of Social and Personal Relationships*, 16(6), 751-771.
- Martin, C. L., & Halverson Jr, C. F. (1981). A schematic processing model of sex typing and stereotyping in children. *Child Development*, 52(4), 1119-1134.
- Martin, C. L., & Little, J. K. (1990). The relation of gender understanding to children's sex-typed preferences and gender stereotypes. *Child Development*, 61(5), 1427-1439.
- Martin, C. L., Ruble, D. N., & Szkrybalo, J. (2002). Cognitive theories of early gender development. *Psychological Bulletin*, 128(6), 903-933.
- Meyer, M., & Gelman, S. A. (2016). Gender Essentialism in Children and Parents: Implications for the Development of Gender Stereotyping and Gender-Typed Preferences. *Sex Roles*, 75(9-10), 409-421.
- Mulvey, K. L., Rizzo, M. T., & Killen, M. (2016). Challenging gender stereotypes: Theory of mind and peer group dynamics. *Developmental Science*, 19(6), 999-1010.

- Powlisha, K. K. (1995). Intergroup processes in childhood: Social categorization and sex role development. *Developmental Psychology, 31*(5), 781-788.
- Rhodes, M., & Gelman, S. A. (2009). A developmental examination of the conceptual structure of animal, artifact, and human social categories across two cultural contexts. *Cognitive Psychology, 59*(3), 244-274.
- Rholes, W. S., & Ruble, D. N. (1984). Children's understanding of dispositional characteristics of others. *Child Development, 55*(2), 550-560.
- Ross, H. S., Recchia, H. E., & Carpendale, J. I. (2005). Making sense of divergent interpretations of conflict and developing an interpretive understanding of mind. *Journal of Cognition and Development, 6*(4), 571-592.
- Ruble, D. N., Martin, C. L., & Berenbaum, S. A. (1998). Gender development. In W. Damon (Series Ed.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (pp. 933-1016). New York: Wiley.
- Serbin, L. A., & Sprafkin, C. (1986). The salience of gender and the process of sex typing in three-to seven-year-old children. *Child Development, 57*(5), 1188-1199.
- Shutts, K. (2013). Is gender “special”? In M. R. Banaji & S. A. Gelman (Eds.), *Navigating the social world: The early years* (pp. 297-300). New York: Oxford University Press.
- Shutts, K., Roben, C. K. P., & Spelke, E. S. (2013). Children's use of social categories in thinking about people and social relationships. *Journal of Cognition and Development, 14*(1), 35-62.

- Signorella, M. L., Bigler, R. S., & Liben, L. S. (1993). Developmental differences in children's gender schemata about others: A meta-analytic review. *Developmental Review, 13*(2), 147-183.
- Signorella, M. L., & Liben, L. S. (1984). Recall and reconstruction of gender-related pictures: Effects of attitude, task difficulty, and age. *Child Development, 55*(2), 393-405.
- Taylor, M. G., Rhodes, M., & Gelman, S. A. (2009). Boys will be boys; cows will be cows: Children's essentialist reasoning about gender categories and animal species. *Child Development, 80*(2), 461-481.
- Trautner, H. M., Ruble, D. N., Cyphers, L., Kirsten, B., Behrendt, R., & Hartmann, P. (2005). Rigidity and flexibility of gender stereotypes in childhood: Developmental or differential? *Infant and Child Development, 14*(4), 365-381.
- Weisgram, E. S., Bigler, R. S., & Liben, L. S. (2010). Gender, values, and occupational interests among children, adolescents, and adults. *Child Development, 81*(3), 778-796.
- Weisman, K., Johnson, M. V., & Shutts, K. (2015). Young children's automatic encoding of social categories. *Developmental Science, 18*(6), 1036-1043.
- Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development, 75*(2), 523-541.
- Yee, M., & Brown, R. (1994). The development of gender differentiation in young children. *British Journal of Social Psychology, 33*(2), 183-196.

## APPENDIX A

### TABLES

Table 1

Descriptive Statistics and Tests Against Chance for Matching in Baseline Condition by Age Group

Age Group	Novel Activities		Friendship	
	<i>M (SD)</i>	Tests against chance	<i>M (SD)</i>	Tests against chance
5-6	0.84 (0.37)	$t(24) = 4.54, p < .001$	0.80 (0.41)	$t(24) = 3.67, p = .001$
7-8	0.96 (0.20)	$t(23) = 11.00, p < .001$	0.92 (0.28)	$t(23) = 7.23, p < .001$

Note. Dependent variable scores: 0 = target matched with stereotypical character, 1 = target matched with ambiguous character (shared appearance and conceptual information).



Table 2

Frequency of Responses for Ambiguity Check Questions for the Unlabeled Condition

Question	Trait			Occupation		
	Boy	Girl	Not Sure	Boy	Girl	Not Sure
1	52.8%	15.1%	32.1%	69.8%	11.3%	18.9%
2	60.4%	15.1%	24.5%	67.9%	9.4%	22.6%
3	32.1%	3.8%	64.2%	30.2%	5.7%	64.2%

Note. Question 1: Do you think this person [point to target] is a boy, a girl, or you're not sure?" Question 2: "Do you think this person [point to target] looks like a boy, a girl, or you're not sure?" Question 3: "Is this person friends with boys, girls, or both?" Note that for the third question, "both" equates to "not sure" on the table above.

Table 3

Frequency of Responses for Ambiguity Check Questions for the Labeled Condition

Question	Trait			Occupation		
	Boy	Girl	Not Sure	Boy	Girl	Not Sure
1	24.5%	9.6%	64.2%	34.0%	7.5%	56.6%
2	26.4%	11.3%	60.4%	28.3%	9.4%	60.4%
3	15.1%	7.5%	75.0%	17.0%	3.8%	77.4%

Note. Question 1: Do you think this person [point to target] is a boy, a girl, or you're not sure?" Question 2: "Do you think this person [point to target] looks like a boy, a girl, or you're not sure?" Question 3: "Is this person friends with boys, girls, or both?" Note that for the third question, "both" equates to "not sure" on the table above.

Table 4

Means, Standard Deviations, and Tests Against Chance for Matching in the Labeled Condition by Age Group and Story Type

		Trait				Occupation			
		Novel Activities		Friendship		Novel Activities		Friendship	
Age	<i>n</i>	<i>M</i> ( <i>SD</i> )	<i>n</i>	<i>M</i> ( <i>SD</i> )	<i>n</i>	<i>M</i> ( <i>SD</i> )	<i>n</i>	<i>M</i> ( <i>SD</i> )	
5-6	27	0.48 (0.51)	27	0.59 (0.50)	27	0.56 (0.51)	27	0.59 (0.50)	
7-8	26	0.65 (0.49)	26	0.73 (0.45)**	26	0.54 (0.51)	26	0.58 (0.50)	

Note. \*\* indicates significance against chance  $p \leq .01$ . Dependent variable scores: 0 = target matched with ambiguous character (shared appearance information), 1 = target matched with stereotypical character (shared conceptual information).

Table 5

Means, Standard Deviations, and Tests Against Chance for Matching in the Unlabeled Condition by Age Group and Story Type

	Trait				Occupation			
	Novel Activities		Friendship		Novel Activities		Friendship	
	<i>n</i>	<i>M(SD)</i>	<i>n</i>	<i>M(SD)</i>	<i>n</i>	<i>M(SD)</i>	<i>n</i>	<i>M(SD)</i>
5-6	26	0.46 (0.51)	26	0.62 (0.50)	26	0.69 (0.47)*	26	0.65 (0.49)
7-8	27	0.74 (0.45)**	27	0.74 (0.45)**	27	0.63 (0.49)	27	0.70 (0.47)*

Note. \*indicates significance against chance at  $p \leq .05$ , \*\* indicates significance against chance  $p \leq .01$ . Dependent variable scores: 0 = target matched with ambiguous character (shared appearance information), 1 = target matched with stereotypical character (shared conceptual information).

Table 6

## Correlations Between Matching Variables and Essentialism Measure

Variable	Question 1	Question 2	Question 3	Question 4	Total
Labeled, Age Group	-0.29	-0.01	-0.16	0.26	-0.07
Labeled, NA	-0.04	0.02	-0.01	-0.05	-0.03
Labeled, Friendship	0.15	0.12	-0.16	-0.10	-0.02
Unlabeled, Age Group	-0.25	0.06	-0.19	0.32	-0.01
Unlabeled, NA	-0.15	-0.11	0.05	0.18	-0.01
Unlabeled, Friendship	0.00	-0.01	-0.16	-0.06	-0.10
Age Group	-0.26**	-0.01	-0.17	0.29**	-0.04
NA	-0.11	-0.06	0.01	0.07	-0.03
Friendship	-0.07	-0.04	-0.17	-0.08	-0.06

Note. \*\* indicates  $p \leq .01$ . NA = Novel Activities Question, Friendship = Friendship Question.

Table 7

Correlations Between Matching Variables and Doodles Task

Variable	Irrelevant-Past	Relevant-Past	Distinct-Pasts	Total
Age Group	0.05	0.21	0.09	0.17
Total NA	0.03	0.06	-0.10	-0.00
Total Friendship	-0.14	-0.07	-0.01	-0.12

Note. \*indicates significance at  $p \leq .05$ . NA = Novel Activities Question, Friendship = Friendship Question.

## APPENDIX B

### FIGURES

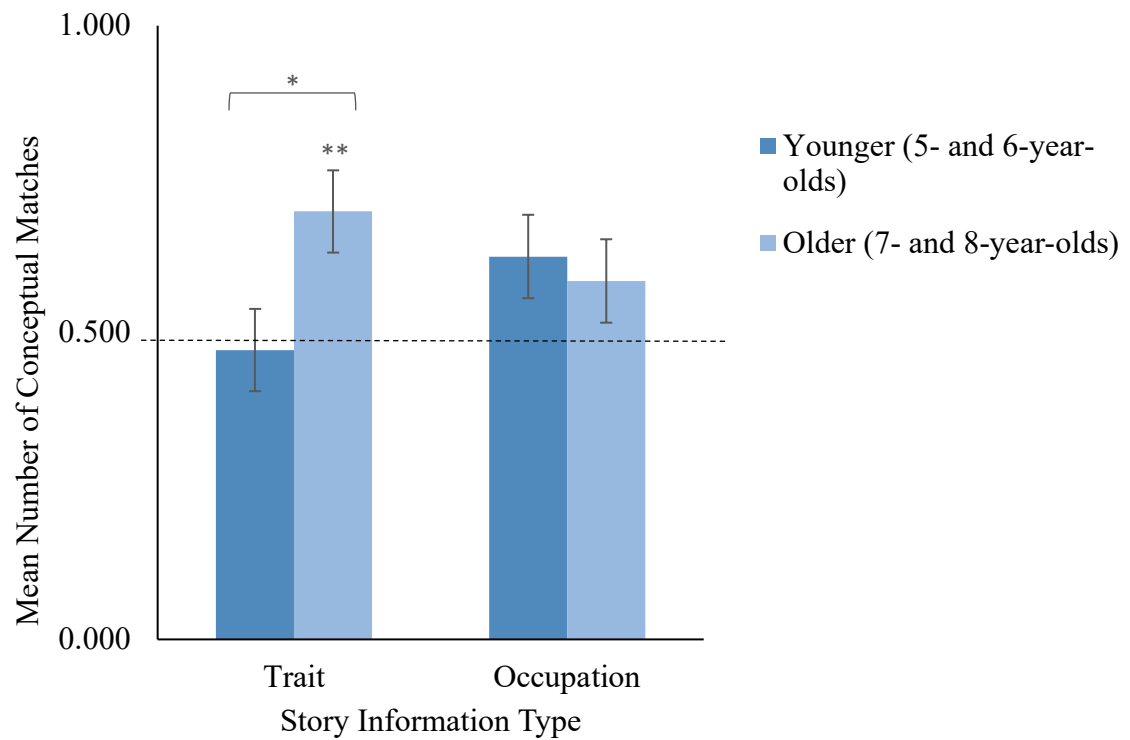


Figure 1. Story Information Type x Age Group Interaction. Mean number of conceptual matches for the Novel Activities Question by Story Information Type and Age Group. Participants received a score of 0 for appearance matches and a score of 1 for conceptual matches on the Novel Activities Question. Scores ranged from 0 – 1. Error bars indicate standard errors. \* indicates  $p \leq .05$ . \*\* indicates significance against chance,  $p \leq .01$ .

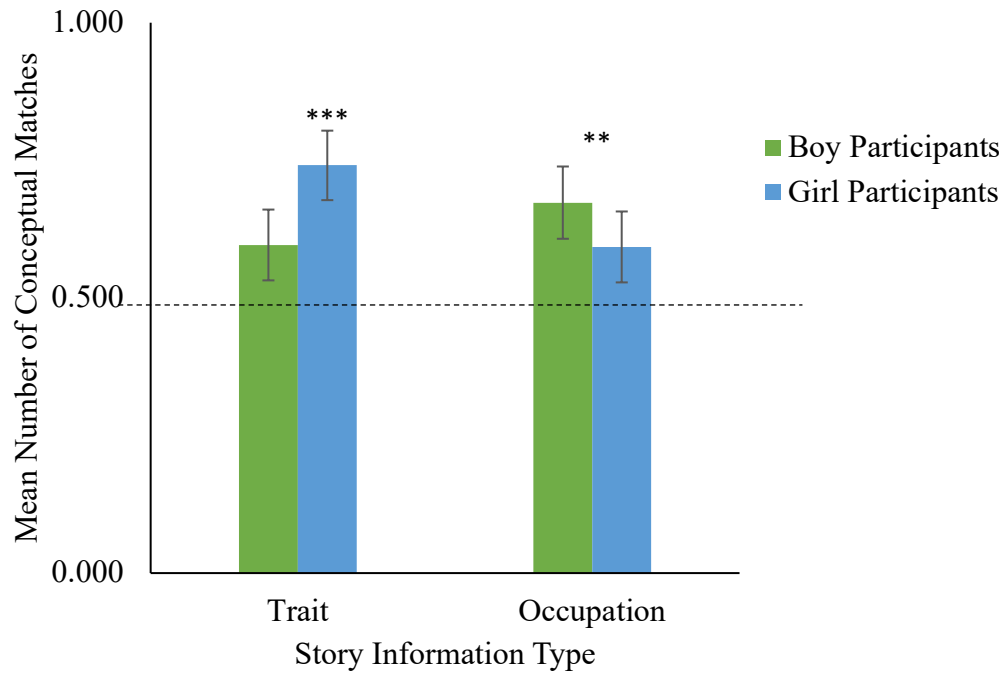


Figure 2. Story Information Type x Participant Gender Interaction. Mean number of conceptual matches for the Friendship Question by Story Information Type and Participant Gender. Participants received a score of 0 for appearance matches and a score of 1 for conceptual matches on the Friendship Question. Scores ranged from 0 – 1. Error bars indicate standard errors. \*\* indicates significance against chance,  $p \leq .01$  and \*\*\* indicates significance against chance,  $p \leq .001$ .



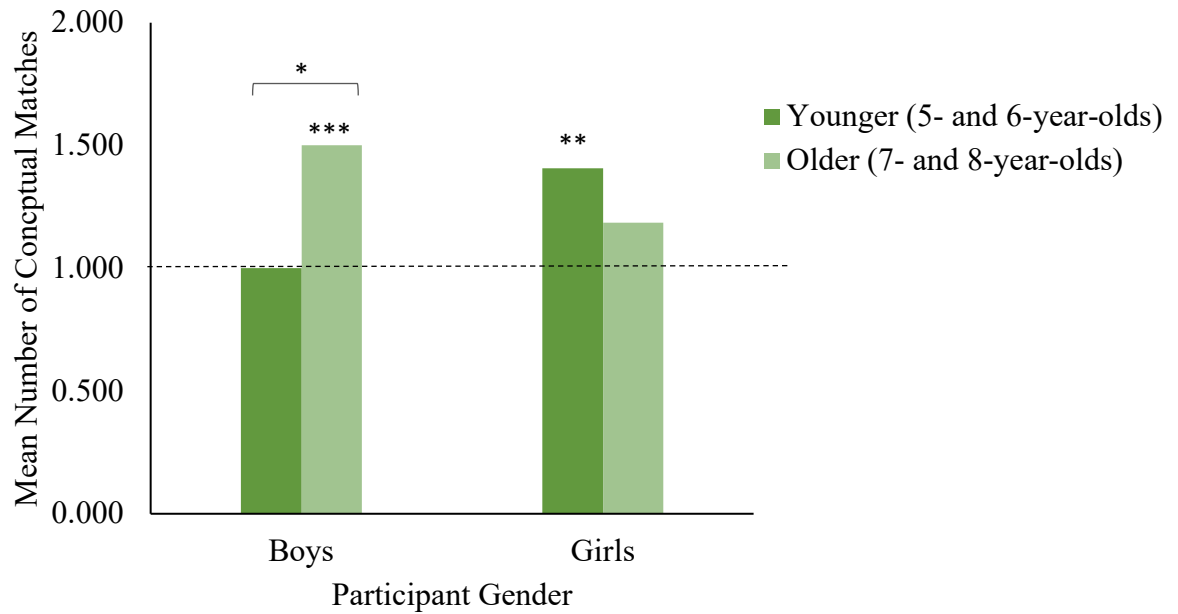
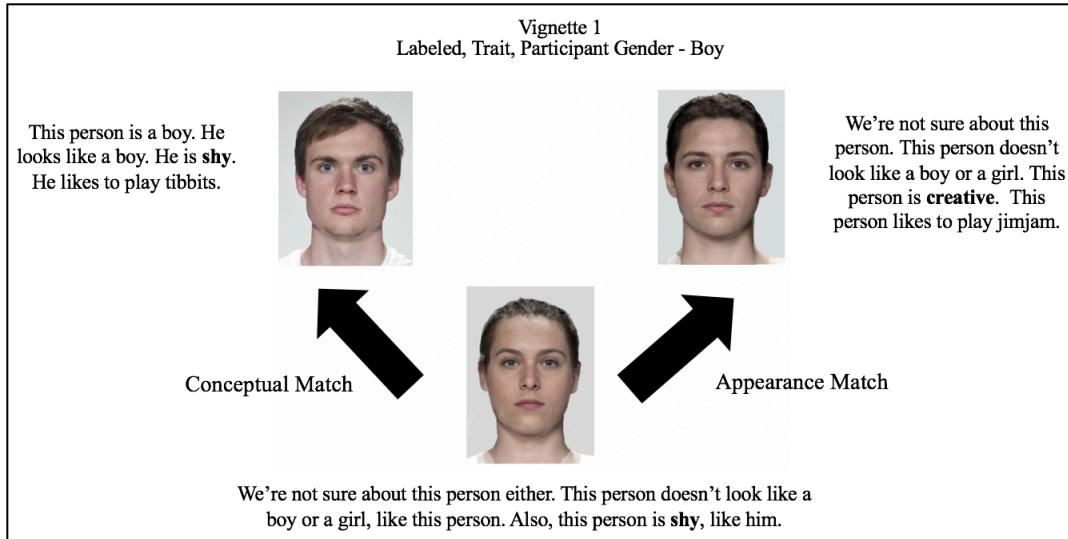


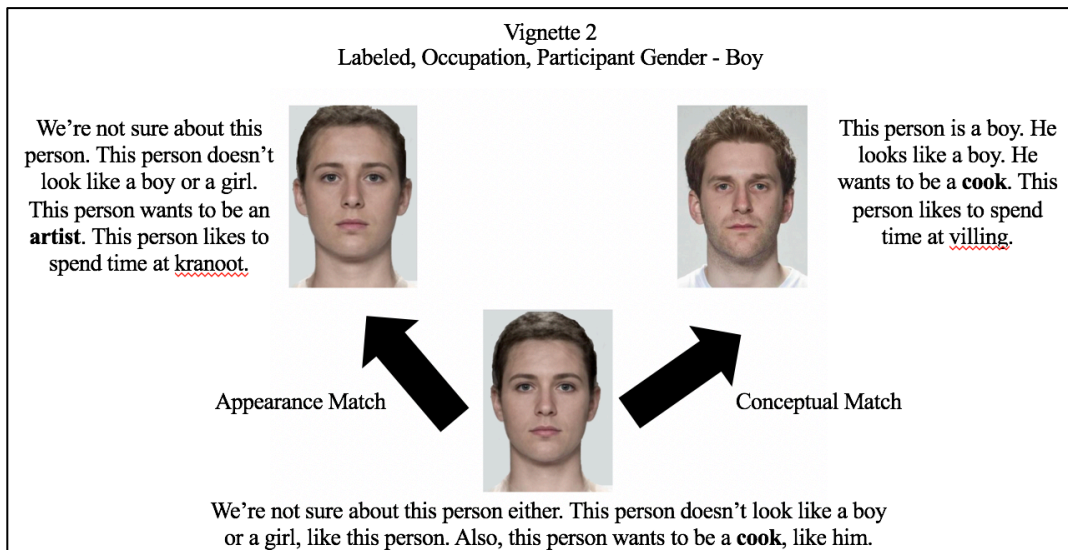
Figure 3. Participant Gender x Age Group Interaction. Mean number of conceptual matches for the Friendship Question by Participant Gender and Age Group. Participants received a score of 0 for appearance matches and a score of 1 for conceptual matches on the Friendship Question. Scores ranged from 0 – 2, as matches were combined across story information type. Error bars indicate standard errors. \* indicates  $p \leq .05$ . and \*\* indicates significance against chance,  $p \leq .01$  and \*\*\* indicates significance against chance,  $p \leq .001$ .

## APPENDIX C

### STIMULI AND VIGNETTES



Sample set, first trio (Labeled, **Trait**, Participant Gender - Boy): Gender stereotypical character - top left, perceptually gender ambiguous character – top right, target – center.



Sample set, second trio (Labeled, **Occupation**, Participant Gender - Boy): Perceptually gender ambiguous character – top left, boy stereotypical character - top right, target – center.

**Labeled Girl, Occupation:** “Here are three people. This person *[point to woman]* is a girl. She looks like a girl. She *[continue pointing]* wants to be a **cook**. She [likes to play jimjam/tibbits OR spend time at villing/kranoot]. We’re not sure about this person *[point to gender ambiguous character]*. This person doesn’t look like a boy or a girl. This person *[continue pointing]* wants to be an artist. This person [likes to play jimjam/tibbits OR spend time at villing/kranoot]. We’re not sure about this person *[point to target]* either. This person doesn’t look like a boy or a girl, like this person *[point to gender ambiguous character]*. Also, this person *[continue pointing to target]* wants to be a **cook**, like her *[point to woman]*.”

**Labeled Boy, Occupation:** “Here are three people. This person *[point to man]* is a boy. He looks like a boy. He wants *[continue pointing]* to be a **cook**. He [likes to play jimjam/tibbits OR spend time at villing/kranoot]. We’re not sure about this person *[point to gender ambiguous character]*. This person doesn’t look like a boy or a girl. This person *[continue pointing]* wants to be an artist. This person [likes to play jimjam/tibbits OR spend time at villing/kranoot]. We’re not sure about this person *[point to target]* either. This person doesn’t look like a boy or a girl, like this person *[point to gender ambiguous character]*. Also, this person *[continue pointing to target]* wants to be a **cook**, like him *[point to man]*.”

**Unlabeled Girls/Boys, Occupation:** “Here are three people. This person *[point to man or woman]* is short. This person wears small clothes. This person *[continue pointing]* wants to be a **cook**. This person [likes to play jimjam/tibbits OR spend time at villing/kranoot]. We’re not sure if this person *[point to gender ambiguous character]* is tall or short. We’re not sure if they wear big or small clothes. This person *[continue pointing]* wants to be an artist. This person [likes to play jimjam/tibbits OR spend time at villing/kranoot]. We’re not sure if this person *[point to target]* is tall or short either. We’re not sure if they wear big or small clothes, like this person *[point to gender ambiguous character]*. This person *[continue pointing to target]* wants to be a **cook**, like this person *[point to man or woman]*.”

**Labeled Girls, Trait:** “Here are three people. This person *[point to woman]* is a girl. She looks like a girl. She is **shy**. She [likes to play jimjam/tibbits OR spend time at villing/kranoot]. We’re not sure about this person *[point to gender ambiguous character]*. This person doesn’t look like a boy or a girl. This person *[continue pointing]* is creative. This person [likes to play jimjam/tibbits OR spend time at villing/kranoot]. We’re not sure about this person *[point to target]* either. This person doesn’t look like a boy or a girl, like this person *[point to gender ambiguous character]*. Also, this person *[continue pointing to target]* is **shy**, like her *[point to woman]*.”

**Labeled Boys, Trait:** “Here are three people. This person *[point to man]* is a boy. He looks like a boy. He is **shy**. He [likes to play jimjam/tibbits OR spend time at villing/kranoot]. We’re not sure about this person *[point to gender ambiguous character]*.”

This person doesn't look like a boy or a girl. This person *[continue pointing]* is creative. This person [likes to play jimjam/tibbits OR spend time at villing/kranoot]. We're not sure about this person *[point to target]* either. This person doesn't look like a boy or a girl, like this person *[point to gender ambiguous character]*. Also, this person *[continue pointing to target]* is **shy**, like him *[point to man]*."

**Unlabeled Girls/Boys, Trait:** "Here are three people. This person *[point to woman]* is short. This person wears small clothes. This person is **shy**. This person [likes to play jimjam/tibbits OR spend time at villing/kranoot]. We're not sure if this person *[point to gender ambiguous character]* is tall or short. We're not sure they wear big or small clothes. This person *[continue pointing]* is creative. This person [likes to play jimjam/tibbits OR spend time at villing/kranoot]. We're not sure if this person *[point to target]* is tall or short either. We're not sure if they wear big or small clothes, like this person. This person *[continue pointing to target]* is **shy**, like this person *[point to woman]*."

**Baseline, Trait:** "Here are three people. This person *[point to man or woman]* is short. This person wears small clothes. This person *[continue pointing]* is creative. This person [likes to play jimjam/tibbits OR spend time at villing/kranoot]. We're not sure if this person *[point to gender ambiguous person]* is tall or short. We're not sure if they wear big or small clothes. This person *[continue pointing]* is **shy**. This person [likes to play jimjam/tibbits OR spend time at villing/kranoot]. We're not sure if this person *[point to target]* is tall or short either. We're not sure if they wear big or small clothes, like this person *[point to gender ambiguous character]*. This person *[continue pointing to target]* is **shy**, like this person *[point to gender ambiguous character]*."

**Baseline, Occupation:** "Here are three people. This person *[point to man or woman]* is short. This person wears small clothes. This person *[continue pointing]* wants to be an artist. This person [likes to play jimjam/tibbits OR spend time at villing/kranoot]. We're not sure if this person *[point to gender ambiguous person]* is tall or short. We're not sure if they wear big or small clothes. This person *[continue pointing]* wants to be a **cook**. This person [likes to play jimjam/tibbits OR spend time at villing/kranoot]. We're not sure if this person *[point to target]* is tall or short either. We're not sure if they wear big or small clothes, like this person. This person *[continue pointing to target]* wants to be a **cook**, like this person *[point to gender ambiguous character]*."

## APPENDIX D

### MEMORY CHECK

*Repeat for each character in triad. Randomize order that characters are asked about (e.g., 1. target, 2. stereotypical character, 3. gender ambiguous character). Randomize gender presentation (labeled condition)/height information (unlabeled condition) for questions 1 and 2 for each character. If participant answers a question incorrectly, re-state information for character. Maximum 3 repeats of information for each question. Exclude participant if more repetitions are needed. Record all answers on scoresheet, including number of repetitions (if necessary).*

#### **Labeled/Occupation, boys & girls:**

- 1) "Did I say this person is a boy, girl, or we're not sure?"
- 2) "Did I say this person looks like a boy, girl, or we're not sure?"
- 3) "Did I say this person wants to be a cook or an artist?"
- 4) Ask for all characters EXCEPT for target: "Did I say this person likes to [play tibbits or jimjam OR spend time at villing/kranoot]?"

#### **Unlabeled/Occupation, boys & girls:**

- 1) "Did I say this person is short, tall, or we're not sure?"
- 2) "Did I say this person wears small clothes, big clothes, or we're not sure?"
- 3) "Did I say this person wants to be a cook or an artist?"
- 4) Ask for all characters EXCEPT for target: "Did I say this person likes to [play tibbits or jimjam OR spend time at villing/kranoot]?"

#### **Labeled/Trait, boys & girls:**

- 1) "Did I say this person is a boy, girl, or we're not sure?"
- 2) "Did I say this person looks like a boy, girl, or we're not sure?"
- 3) "Did I say this person is shy or creative?"
- 4) Ask for all characters EXCEPT for target: "Did I say this person likes to [play tibbits or jimjam OR spend time at villing/kranoot]?"

#### **Unlabeled/Trait, boys & girls:**

- 1) "Did I say this person is short, tall, or we're not sure?"
- 2) "Did I say this person wears small clothes, big clothes, or we're not sure?"
- 3) "Did I say this person is shy or creative?"
- 4) Ask for all characters EXCEPT for target: "Did I say this person likes to [play tibbits or jimjam OR spend time at villing/kranoot]?"

#### **Baseline, occupation:**

- 1) "Did I say this person is short, tall, or we're not sure?"

- 2) “Did I say this person wears small clothes, big clothes, or we’re not sure?”
- 3) “Did I say this person wants to be a cook or an artist?”
- 4) Ask for all characters EXCEPT for target: “Did I say this person likes to [play tibbits or jimjam OR spend time at villing/kranoot]?”

**Baseline, trait:**

- 1) “Did I say this person is short, tall, or we’re not sure?”
- 2) “Did I say this person wears small clothes, big clothes, or we’re not sure?”
- 3) “Did I say this person is shy or creative?”
- 4) Ask for all characters EXCEPT for target: “Did I say this person likes to [play tibbits or jimjam OR spend time at villing/kranoot]?”

## APPENDIX E

### DEPENDENT MEASURES

#### Novel Activities Question:

***Only administer for target. Randomize order of character information presented within the question (e.g., 1. ambiguous character information, 2. stereotypical character information). Record answers on scoresheet.***

- a) Trait story: “Does this person like to [play **TIBBITS** / **JIMJAM OR** spend time at **VILLING** / **KRANOOT**] like this person who is **SHY** / **CREATIVE** or does this person like to [play **TIBBITS** / **JIMJAM OR** spend time at **VILLING** / **KRANOOT**] like this person who is **SHY** / **CREATIVE**?”
- b) Occupation story: “Does this person like to [play **TIBBITS** / **JIMJAM OR** spend time at **VILLING** / **KRANOOT**] like this person who wants to be a **COOK** / **ARTIST** or does this person like to [play **TIBBITS** / **JIMJAM OR** spend time at **VILLING** / **KRANOOT**] like this person who wants to be a **COOK** / **ARTIST**?”

#### Friendship Question:

***Only administer for target. Record answers on scoresheet.***

- a) “Who should this person [point to target] be friends with? Why?”

#### Target character activity inferences:

***Only administer for the target. Record answers on scoresheet.***

- 1) Does this person like to play with dolls\*, toy trucks, or both?
- 2) Does this person like movies about princesses\*, soldiers, or both?
- 3) Does this person know a lot about ballet\*, football, or both?
- 4) Does this person like to build with tools, bake cookies\*, or both?
- 5) Does this person like to sew\*, **go to the beach**, or both?
- 6) Does this person like to play video games, **go to the movies**, or both?
- 7) Does this person want to become an auto mechanic, a teacher\*, or both?
- 8) Does this person want to become a firefighter, a nurse\*, or both?
- 9) Does this person want to become an engineer, a **baker**, or both?
- 10) Does this person want to become a secretary\*, a **writer**, or both?

\*=stereotypical for girls; **bold** = gender neutral; all others = stereotypical for boy

#### Character liking, affiliation, and trait attribution

***Administer for each character. Randomize order that each question is presented.***

***Answer recorded on scoresheet.***

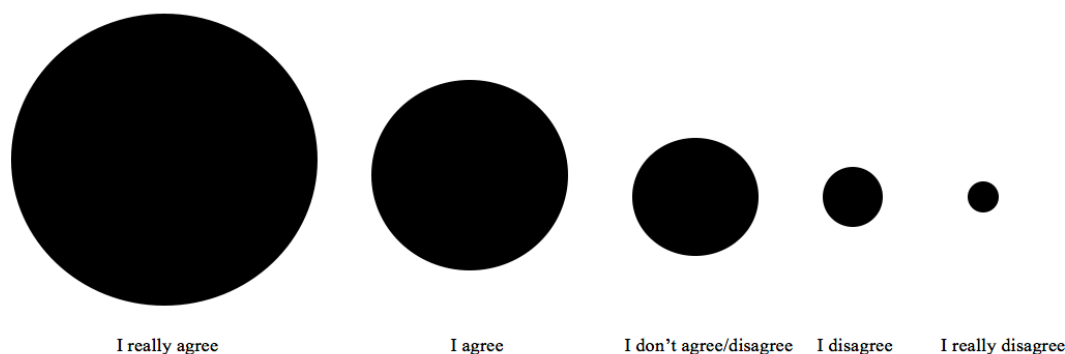
- a) Liking: “How much do *you* like this person [point to target] – a lot, a little, or in the middle?”
- b) Affiliation: “How much would *you* like to be friends with this person [point to target] – a lot, a little, or in the middle?”
- c) Trait attribution: “Do you think this person [point to target] is nice, mean, or not nice or mean?”

APPENDIX F

ESSENTIALISM MEASURE

*Administer training prior to questions. Record answers on answer sheet.*

**Scale:**



**Training:**

- 1) “For this next question, we’re going to use these circles. This one over here [point to largest circle – on the left] is *really big* and means “I really agree” This one over here is *just big* [point to second largest circle] and means “I agree” and this one in the middle [point to middle circle] is *kind of big* and means “I don’t agree/disagree” This next one is *not really big* [point to second smallest circle] and means “I disagree” and this one all the way over here [point to smallest circle – on the right] is *not big at all* – it’s tiny – and it means “I really disagree.”
- 2) “Let’s do some practice questions. What’s something you like to do *a whole lot*?...So where would you point to show how much you agree if I said “You like \_\_\_\_.” And what’s something you *do not like at all*?...So where would you point to show how much you agree if I said “You like \_\_\_\_.” And what’s something *you sort of like, you’re in the middle about it*?...So where would you point to show how much you agree if I said “You like \_\_\_\_.” Okay, now we can ask some new questions.”

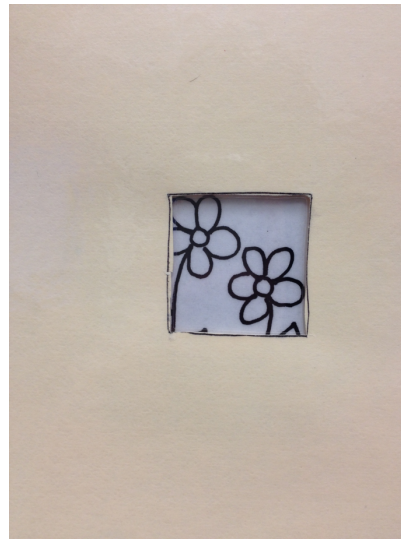
**Essentialism questions:**

- 1) “Being a boy or girl is a very important part of what makes kids who they are.”
- 2) “Boys have many things in common with boys and girls have many things in common with girls.”
- 3) “Knowing if a kid is a girl or a boy tells you a lot about the kid.”
- 4) “Kids are either boys or girls, there is nothing in between.”

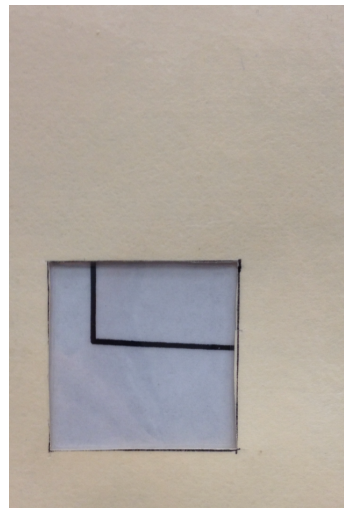
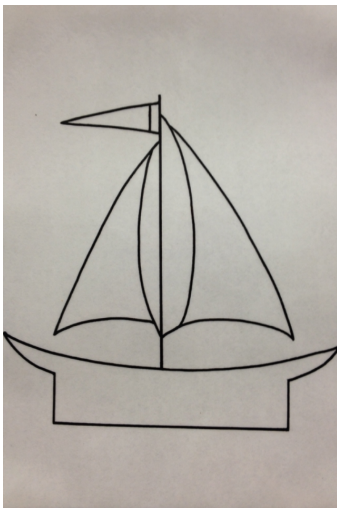


## APPENDIX G

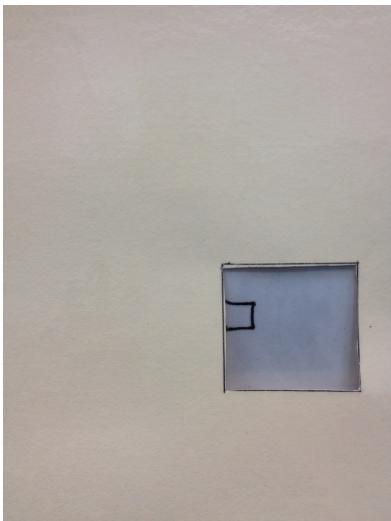
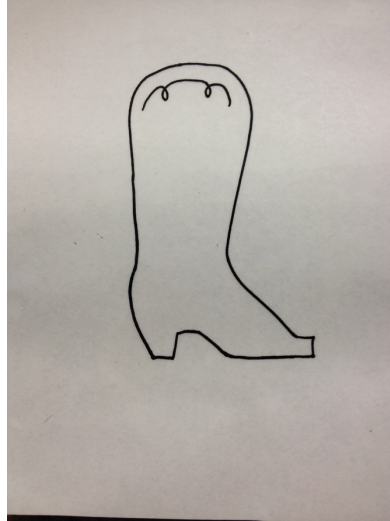
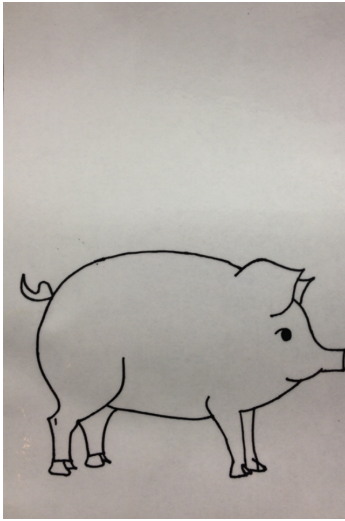
### DROODLES TASK (FROM LAGATTUTA ET AL., 2010)



**Irrelevant-Past Trial:** Distinguishing part of drawing is visible. All characters (including naïve observer) are knowledgeable about drawing content, regardless of whether they saw the content of the drawing prior to the placement of the occluder.



**Relevant-Past Trial:** Only ambiguous portion of drawing is visible. Only the character that saw the boat prior to the placement of the occluder is knowledgeable about the content of the drawing. Naïve character does not know the content of the drawing.



**Distinct-Past Trial:** Only ambiguous portion of drawing is visible. Each character saw a different picture (i.e., only pig or only boot). Only the character that saw the pig prior to the placement of the occluder is knowledgeable about the content of the drawing. Naïve character does not know the content of the drawing.